



NIH PUBLIC ACCESS

Author Manuscript

Qual Life Res. Author manuscript; available in PMC 2013 October 07.

Published in final edited form as:

Qual Life Res. 2012 September ; 21(7): 1223–1240. doi:10.1007/s11136-011-0025-2.

PROMIS® Parent Proxy Report Scales: an item response theory analysis of the parent proxy report item banks

James W. Varni,

Department of Pediatrics, College of Medicine, Department of Landscape Architecture and Urban Planning, College of Architecture, Texas A&M University, 3137 TAMU, College Station, TX 77843-3137, USA

David Thissen,

Department of Psychology, University of North Carolina, at Chapel Hill, Chapel Hill, NC, USA

Brian D. Stucky,

RAND Corporation, Santa Monica, CA, USA

Yang Liu,

Department of Psychology, University of North Carolina, at Chapel Hill, Chapel Hill, NC, USA

Hally Gorder,

Department of Psychology, University of North Carolina, at Chapel Hill, Chapel Hill, NC, USA

Debra E. Irwin,

Department of Epidemiology, University of North Carolina, at Chapel Hill, Chapel Hill, NC, USA

Esi Morgan DeWitt,

Department of Pediatrics, Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Jin-Shei Lai,

Department of Medical Social Sciences, Northwestern, University Feinberg School of Medicine, Chicago, IL, USA

Dagmar Amtmann, and

Department of Rehabilitation Medicine, University of Washington, Seattle, WA, USA

Darren A. DeWalt

Division of General Medicine and Clinical Epidemiology, Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Abstract

Objective—The objective of the present study is to describe the item response theory (IRT) analysis of the National Institutes of Health (NIH) Patient Reported Out-comes Measurement Information System (PROMIS®) pediatric parent proxy-report item banks and the measurement properties of the new PROMIS® Parent Proxy Report Scales for ages 8–17 years.

Methods—Parent proxy-report items were written to parallel the pediatric self-report items. Test forms containing the items were completed by 1,548 parent–child pairs. CCFA and IRT analyses of scale dimensionality and item local dependence, and IRT analyses of differential item functioning were conducted.

Results—Parent proxy-report item banks were developed and IRT parameters are provided. The recommended unidimensional short forms for the PROMIS® Parent Proxy Report Scales are item sets that are subsets of the pediatric self-report short forms, setting aside items for which parent responses exhibit local dependence. Parent proxy-report demonstrated moderate to low agreement with pediatric self-report.

Conclusions—The study provides initial calibrations of the PROMIS® parent proxy-report item banks and the creation of the PROMIS® Parent Proxy-Report Scales. It is anticipated that these new scales will have application for pediatric populations in which pediatric self-report is not feasible.

Keywords

PROMIS®; Parent proxy report; Item response theory

Introduction

The Patient Reported Outcomes Measurement Information System (PROMIS®) is a National Institutes of Health (NIH) Initiative, created to advance the assessment of patient-reported outcomes (PRO) in chronic diseases. Items are evaluated using item response theory (IRT) to derive scales with scores that are maximally reliable and valid along the full spectrum of the latent trait [1]. A primary objective is to develop item banks and computerized adaptive tests (CAT) across a variety of chronic disorders [2]. During the past 7 years, the PROMIS® Pediatric Cooperative Group has developed pediatric self-report item banks for ages 8–17 years across five generic health domains (physical function, pain, fatigue, emotional health, and social health) consistent with the larger PROMIS® network [3]. It was anticipated that measures of these five generic health domains would be applicable across pediatric chronic health conditions and hence were developed as generic or non-disease-specific scales [4-7]. Additionally, an asthma-specific item bank was developed and validated [8].

It has been well documented in both the adult and pediatric literature that information provided by proxy-respondents is not equivalent to that reported by the patient [9, 10]. Imperfect agreement between self-report and proxy-report, termed cross-informant variance [11], has been consistently documented in the HRQOL measurement of children with chronic health conditions and healthy children [12]. However, even as pediatric patient self-report is advocated, there remains a role for parent proxy-report in pediatric clinical trials and health services research.

While pediatric patient self-report should be considered the standard for measuring PROs, there may be circumstances when the child is too young, too cognitively impaired, too ill or fatigued to complete a PRO instrument, and parent proxy-report may be needed in such cases [13]. Further, it is typically parents' perceptions of their children's health and well-being that influences healthcare utilization [14-16]. Thus, instruments should be developed that measure the perspectives of both the child/adolescent and parent since these perspectives may be independently related to healthcare utilization, risk factors, and quality of care [17].

The majority of parent proxy-report scales, consistent with other clinical assessment instruments [18], have utilized classical test theory (CTT) and have rarely taken advantage of IRT analysis in the scale development process [19]. By utilizing 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 [19].

The objective of the present study is to address this measurement gap in the parent proxy-report literature by describing the IRT analysis of the PROMIS® parent proxy-report item banks and the measurement properties of the new PROMIS® Parent Proxy Report Scales, including investigations of scale dimensionality and sources of local dependence and differential item functioning.

Methods

Participants

Participants were recruited between May 2008 through March 2009 in hospital-based outpatient general pediatrics and subspecialty clinics. Clinic participants were identified through a review of clinic appointment rosters or while in the clinic waiting rooms according to protocols approved by the institutional review boards (IRBs) of University of North Carolina (UNC), Duke University Medical Center, University of Washington (UW), Children's Memorial Hospital, Chicago (CMH), and Children's Hospital at Scott and White (S&W) in Texas. Pediatric patients within the appropriate age range who had clinic appointments and their caregivers were recruited while waiting for their clinic appointments. The UNC, Duke, UW, CMH and S&W general pediatric clinics were representative of health issues for which children have physician office visits (e.g., well child visits, acute illnesses, and some chronic illnesses). The specialty clinics included Pulmonology, Allergy, Gastroenterology, Rheumatology, Nephrology, Obesity, Rehabilitation, Dermatology, and Endocrinology. Children with asthma were over sampled during recruitment because asthma-specific items were tested.

To be eligible to participate in the large-scale testing survey, all participants were required to meet the following inclusion criteria: able to speak and read English; and able to see and interact with a computer screen, keyboard, and mouse. Children enrolled were between the ages of 8 and 17 years and with their parents/guardians formed a dyad (for convenience, we refer to the dyad as simply parent/child). Both members of the dyad were required to individually complete the items. Children completed the self-report version, and parents completed the proxy-report version.

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 the measures implemented to protect participant privacy. Both the informed assent and the informed consent were administered in English, so parents were also required to read and speak English. Each participant 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.

Pediatric self-report item bank development

The PROMIS® Pediatric item banks were developed using a strategic item generation methodology adopted by the PROMIS® Network [2]. 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. The final pediatric self-report item banks contained 165 items across the 5 generic health domains (physical function, pain, fatigue, emotional health, and social health) and Asthma. Because physical function includes both upper extremity and mobility item banks, emotional distress includes separate anger, anxiety

and depressive symptoms item banks, and fatigue includes both fatigue and lack of energy item banks, a total of 10 content domains were tested [4-8].

Parent proxy sampling plan and item distribution

The parent proxy-report items were developed from the 10 existing pediatric self-report content domains [4-8]. The items were revised to retain their meaning, while modifying the phrasing so that all items involved parents reporting on their child. For example, in the pediatric self-report pain interference domain [6], children responded to the item “I had trouble sleeping when I had pain,” while parents responded to the parent proxy-report equivalent of this item, “My child had trouble sleeping when he/she had pain.”

Proxy-report short form items were selected from items that were on the pediatric self-report short forms for each domain and did not include any items that were not already on the self-report short forms. This decision was made because most researchers prefer parent proxy-report item banks that have the same item content as the pediatric self-report item banks.

The 293 proxy-report items from the 10 content domains were administered to 1,548 parents of 8- to 17-year-old children also participating in the study (see Table 1 for sample demographics and Table 2 for content domains). To reduce respondent burden, a multi-form design was used in which the items were divided among nine test forms, and each parent was administered one of the nine forms. To ensure an adequate number of individuals responded to all items, each item appeared on three of the forms. This process resulted in all items being administered to at least 428 parents. A detailed description of the item administration protocol can be found in Irwin et al. [20].

All items had a 7-day recall period and used standardized 5-point response options (e.g., *never, almost never, sometimes, often, almost always*; or, *with no trouble, with a little trouble, with some trouble, with a lot of trouble, not able to do*). Of the 293 items administered, 165 were retained for analysis, as these corresponded to the final items in the pediatric self-report item banks. A complete list of the 165 items may be found in the “Appendix 1”.

Statistical and psychometric methods

The purpose of the psychometric assessment of the parent proxy-report items was to develop parent content domain short forms that could serve as *proxies* for their child counterpart domain short forms. Based on the methods used in the development of the pediatric item banks [4], ten content domain-specific analyses were conducted which investigated scale dimensionality and sources of local dependence (LD).

LD occurs when items retain an association after accounting for inter-item correlations with the latent variable. Researchers typically identify locally dependent item pairs or clusters either using categorical confirmatory factor analysis (CCFA) models that account for LD with correlated residuals and subfactors, or using IRT methods involving model-based pairwise residuals. We employed both methods (CCFA and IRT) to identify potential LD.

Initially, we conducted CCFAs of the inter-item polychoric correlation matrices using *Mplus* [21]. Because this approach is conducted using complete case data, and our sampling design assigned only subsets of items from each content domain to any given parent, multiple CCFAs (usually 2) were conducted on independent subsets of items. This approach involved fitting 15 different CCFA models in which each analysis included at least 175 parents. We began with a single-factor model and estimated additional correlated residuals and subfactors using item content and statistical relevance for guidance. To quantify the degree to which each scale is approximately unidimensional, we used the value of “explained

common variance” (ECV) attributable to the first (general) factor of the final bifactor model for each scale within each combined form [22].

To further investigate potential LD, we calibrated each content domain using Samejima’s graded response model (GRM) in an IRT framework using IRTPRO [23]. Following these initial calibrations, we tested for LD between pairs and clusters of items using a standardized χ^2 statistic as implemented in IRTPRO. The Chen-Thissen χ^2 [24] is based on a comparison between the observed and IRT-expected response frequencies for pairs of items.

These analyses provided two different assessments of LD. Our initial assessment of the validity of these tests of LD was made by comparing the magnitude of the CCFA-based residual correlation to the magnitude of the IRT-based standardized χ^2 statistic. In making these comparisons, we note that the two methods use slightly different respondents; the IRT analyses used the same individuals as the CCFA analyses, and additional parents. Where there was disagreement between the two methods, a team of six content experts evaluated the validity of the potential LD. The expert panel reviewed all pairs and triplets of locally dependent items that contained more than one item from the pediatric short form. If the LD χ^2 statistic was greater than 10, item(s) were eliminated so that only one item from the LD pair or triplet remained. If the LD χ^2 statistic was less than 10, each member of the panel voted on whether or not to delete item(s) and if so, which item to retain. Members made decisions about whether to eliminate item(s) by evaluating whether the item’s content was substantively different from others in the locally dependent group and offered unique information about the child’s experience. When members voted to delete item(s), decisions about which item to retain were based on a content analysis of which item best fit the conceptualization of the domain. Setting aside items from locally dependent pairs ensured that the IRT assumption of unidimensionality was maintained.

We tested for the presence of differential item functioning (DIF) between parents of children ages 8 through 12 and parents of adolescents ages 13 through 17, and separately between parents of male or female children. DIF was investigated using IRTPRO’s DIF module that implements Wald χ^2 tests [25, 26]. In each case, the presence of DIF indicates that the relation between the item responses and latent variable differs between groups. Because many tests of DIF were conducted, we controlled for multiple comparisons using the Benjamini-Hochberg procedure [27]. A significant χ^2 statistic indicates the presence of DIF. To evaluate the influence of DIF on item scores, we used graphical methods suggested by Steinberg and Thissen [28]. The goodness of fit of the IRT model to the data was examined using the S- χ^2 statistic [29] as implemented in IRTPRO. Non-significant S- χ^2 values suggest adequate fit of the model to the data.

In order to place the parent scales on the same metric as the child items, final calibrations of unidimensional domains were conducted using as the reference group mean and variance values (in Table 2) obtained from the child responses who had parents taking the same domain items; parents who did not have children taking the same domain items were included in the calibration as a separate group.

Parent proxy-report short forms were created after setting aside locally dependent items from the item subsets derived from the pediatric self-report short forms. The degree to which these short forms produce precise scores was evaluated using information functions. IRT conceptions of information allow score precision to vary across the range of the content domain. To ease interpretability, reliability is approximately one less the inverse of test information. Thus, when information is 10, reliability is 0.90.

We used two-dimensional IRT models to estimate the correlations between the latent variables measured by the pediatric self-report scales for ages 8–17 and the parent proxy-

report scales for ages 8–17. In these models, one latent variable was used with the graded model to fit the pediatric self-report responses, and a second latent variable was used to fit the parent proxy-report data; the correlation between the two latent variables was estimated simultaneously with the item parameters. In terms used by traditional test theory, this correlation is an estimate of the “disattenuated” correlation between the two variables; that is, the correlation corrected for the presence of measurement error in scores. As a result, these correlations would be 1.0 if the pediatric self-report and parent proxy-report scales measured the same constructs.

Results

Table 3 summarizes the results of the CCFA analyses of dimensionality, tabulating the value of ECV for the general factor for each scale for each form or pair of forms for which the sample size was sufficient. The scales are strongly univocal, with 82–100% of the common variance explained by the general factor. Some degree of local dependence was detectable as statistically significant in these analyses, and the item pairs or clusters involved are noted in the tables in the “Appendix 1”; however, the principle result of the factor analytic examination of the scales was that there is so little residual common variance from a single factor that for practical purposes, unidimensional modeling may proceed.

Tables 4, 5, 6, 7, 8 and 9 in the “Appendix 1” list the items (arranged by the magnitude of the IRT discrimination parameter), IRT parameter estimates, item fit, and DIF statistics for the 165 items retained for analysis and 10 content domains. The notes at the bottom of each table indicate with superscripts items that may exhibit LD and items that were set aside from each domain’s short form. Items marked with the same superscript letter within a domain remain in the item banks, but may exhibit LD, and so users are cautioned to use only one item from each such lettered set in any custom form or CAT. Because many CCFA and IRT models were used to assess the properties of the items, we summarize the results by content domain.

Emotional Distress: LD and DIF

Emotional distress (Table 4) comprises depressive symptoms, anxiety, and anger. No significant DIF by sex or age was identified for these content domains. Among the depressive symptoms short form items, three pairs of items were identified as being locally dependent. From the pairs “My child felt lonely” and “My child felt alone”, and “My child felt sad” and “My child felt unhappy”, the second item was set aside from each LD pair, respectively, for the proxy short form. From the pair “My child felt everything in his/her life went wrong” and “My child felt like he/she couldn’t do anything right”, neither item was set aside after content review. There were three additional pairs of potentially locally dependent items identified in the anxiety domain; however, statistical tests of LD were mixed regarding these pairs, and no further items were set aside from the short form after item content review suggested that significant LD may be spurious.

Short forms were then made for the depressive symptoms (6 items), anxiety (8 items), and anger (5 items) domains using the items from the PROMIS® pediatric self-report short forms, setting aside two items on the depressive symptoms scales that exhibited LD. Figure 1 indicates that the short forms for depressive symptoms and anxiety produce information values greater than 10 from the mean of the latent variable to more than three standard deviations above the mean. The anger scale, containing just 5 items, produces scores with reliability greater than 0.80 from about two standard deviations below the mean to more than three standard deviations above the mean.

Fatigue: LD and DIF

Results for fatigue (the “Lack of Energy” and “Tired” scales; Table 5) indicate two and potentially three locally dependent clusters of items, respectively. After reviewing the content of each LD cluster of items, the review panel concluded that each item provided a unique substantive contribution to the final scale; no items were set aside. Additionally, one item “My child had enough energy to do the things s/he likes to do” had significant DIF with respect to sex, and an additional item “My child needed to sleep during the day” had significant DIF related to age. Though the final short form includes these items, investigators may wish to exclude them if age and sex variables are used to differentiate levels of fatigue. The resulting 8 and 10 item short forms (Lack of Energy and Tired, respectively) both provide reliable scores from about one-half of a standard deviation below the mean to about three deviations above the mean (see Fig. 2).

Physical Functioning: LD and DIF

Results for physical functioning (upper extremity and mobility; Table 6) indicate a single instance of LD between the item pair “My child could do sports and exercise that other kids his/her age could do” and “My child has been physically able to do the activities he/she enjoys most”. The content review panel concluded that each item provides sufficient unique contribution to warrant inclusion in the final short form. Additionally, there was evidence of both age and sex DIF throughout both scales, but this was largely due to missing or sparsely endorsed response categories at the extreme ends of the distribution, and no items were set aside. The resulting 8-item short forms produce reliable scores between about one standard deviation to three standard deviations below the mean (see Fig. 2).

Pain Interference: LD and DIF

There were two potentially locally dependent pairs of items in the pain interference domain, marked with superscripts *a* and *b* in Table 7. All four of these items were considered to be substantively unique, and neither item was set aside from the final short form. Additionally, the item “It was hard for my child to pay attention when he/she had pain” exhibited evidence of DIF by sex; the item was not set aside because it had only 6 and 3 parents of male children endorse the two most extreme response categories, as compared to 20 and 15 parents of female children; hypothesis testing using such sparse cell counts is untrustworthy. The resulting 8-item short form produced scores with information greater than 10 from about one standard deviation below the mean to about two and half standard deviations above the mean (see Fig. 2).

Peer Relationships: LD and DIF

There were two potentially locally dependent pairs of items in the Peer Relationships domain (Table 8). From the item pair “My child was able to count on his/her friends” and “My child was able to talk about everything with his/her friends”, the latter item was set aside. Neither item was set aside from the other LD pair after content review suggested a unique contribution from each item. There were no items with significant DIF. The remaining 7 items on the short form produce scores with reliability greater than 0.90 between the mean and about three standard deviations below the mean (see Fig. 2).

Asthma Impact: LD and DIF

There was no evidence of LD or DIF for the Asthma Impact domain (Table 9). The 8-item short form produced scores with information values greater than 10 between about one standard deviation below the mean to nearly three standard deviations above the mean (see Fig. 2).

Parent/child correlations

The correlations between the latent variables measured by the pediatric self-report scales and the parent proxy-report scales are in Table 2.

Summed scores

For the convenience of users who prefer to use summed scores, scoring tables that translate summed scores on the parent proxy-report short forms into IRT scaled scores are given in Tables 10, 11 [30]. Computation of these scales used the means and standard deviations in Table 2. Because those means and standard deviations refer back to the original calibration samples for the PROMIS® pediatric self-report scales, the effect is that it is as though these parent proxy-report scales had also been calibrated, or “normed”, on that same sample. Because the correlations (in Table 2) between the pediatric self-report scales and the parent proxy-report scales with the same names range from moderate to low, this does not mean that parent proxyreport scores are comparable with pediatric self-report scores. However, it does mean that the average “level” of the scores (on the PROMIS® T-score scales) has the same meaning with respect to “average” and “one standard deviation above average,” and so on.

Discussion

This study describes the development and calibrations of the NIH PROMIS® Parent Proxy Report Scales based on an iterative series of IRT analyses regarding scale dimensionality, item LD, and DIF. After determining scale dimensionality, items with LD and DIF were next identified, and some were removed from the recommended short forms.

The potential advantages of utilizing IRT analysis in scale development include greater flexibility in selecting items from the existing parent proxy-report item banks tailored to the objectives of a particular clinical research investigation. Further, scales that have been developed with CTT may have gaps in their ability to measure the full spectrum of the latent construct, while with IRT calibrated items, one can construct a measure that is useful across the full continuum of the latent variable [19]. Thus, this analytic methodology provides clinical researchers the opportunity to select the most meaningful items for their study design and hypotheses. In this study, we proposed short forms measuring each of the content domains; however, a smaller subset of items from the item banks can also be used and scored on the same metric as the larger set using a more dynamic CAT algorithm.

Our finding that parent proxy-report demonstrated moderate to low agreement with pediatric self-report is consistent with the extant literature [12], suggesting that information provided by proxy-respondents is not equivalent to that reported by the patient. In the HRQOL literature, imperfect agreement between self-report and proxyreport has been consistently documented, typically demonstrating higher correlations for more observable domains (i.e., physical functioning) and lower correlations for less observable or internal symptoms such as emotional functioning, pain, and fatigue [31]. Our findings are consistent with this larger literature.

By administering the items spread over several test forms, we were unable to perform factor analyses across the entire item bank for six of the ten content domains. For each of these six domains, factor analysis was conducted on two separate sets of items. It is possible that factor analyses would turn out differently if all the items within each content domain were analyzed as a single set. However, because the 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 recruited participants from clinics across five sites 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.

Future research with other samples may reveal other sources of DIF for the items; an advantage of IRT as a method is that it can detect item-level DIF, and “flag” items to be used only with caution for comparisons across levels of a variable for which DIF exists. Although analysis of DIF led to smaller item banks, we believe this approach will ultimately yield a more broadly applicable measure for comparing results across populations.

In conclusion, this study provides initial IRT calibrations of the PROMIS® parent proxy-report item banks and the creation of the PROMIS® Parent Proxy Report Scales which address an important gap in the current literature. Further research is indicated on construct validity and tests of the responsiveness of these scales and item banks in larger samples of parents of pediatric patients with chronic health conditions.

Acknowledgments

This work was funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U01AR052181. Information on the Patient-Reported Outcomes Measurement Information System (PROMIS®) can be found at <http://nihroadmap.nih.gov/> and <http://www.nihPROMIS.org>.

Appendix: item parameters and short form scoring tables

See Tables 4, 5, 6, 7, 8, 9, 10 and 11.

Abbreviations

PROMIS®	Patient Reported Outcomes Measurement Information System
FDA	Food and drug administration
HRQOL	Health-related quality of life
NIH	National Institute of Health

References

1. Ader DN. Developing the patient-reported outcomes measurement information system (PROMIS). *Medical Care*. 2007; 45(Suppl 1):S1–S2. [PubMed: 18027399]
2. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-report outcomes measurement information system (PROMIS). *Medical Care*. 2007; 45(Suppl 1):S22–S31. [PubMed: 17443115]
3. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The patient-reported outcomes measurement information system (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. *Medical Care*. 2007; 45(Suppl 1):S3–S11. [PubMed: 17443116]
4. Irwin DE, Stucky, B D, , Thissen D, DeWitt EM, Lai JS, Yeatts K, et al. Sampling plan and patient characteristics of the PROMIS pediatrics large-scale survey. *Quality of Life Research*. 2010; 19:585–594. [PubMed: 20204706]

5. Irwin DE, Stucky BD, Langer MM, Thissen D, DeWitt EM, Lai JS, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Quality of Life Research*. 2010; 19:595–607. [PubMed: 20213516]
6. Varni JW, Stucky BD, Thissen D, DeWitt EM, Irwin DE, Lai JS, et al. PROMIS pediatric pain interference scale: An item response theory analysis of the pediatric pain item bank. *Journal of Pain*. 2010; 11:1109–1119. [PubMed: 20627819]
7. DeWitt EM, Stucky BD, Thissen D, Irwin DE, Langer M, Varni JW, Lai JS, Yeatts KB, DeWalt DA. Construction of the eight-item patient-reported outcomes measurement information system pediatric physical function scales: built using item response theory. *Journal of Clinical Epidemiology*. 2011; 64(7):794–804. [PubMed: 21292444]
8. Yeatts K, Stucky BD, Thissen D, Irwin DE, Varni JW, DeWitt EM, et al. Construction of the pediatric asthma impact scale (PAIS) for the patient-reported outcomes measurement information system (PROMIS). *Journal of Asthma*. 2010; 47:295–302. [PubMed: 20394514]
9. Sprangers MAG, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. *Journal of Clinical Epidemiology*. 1992; 45:743–760. [PubMed: 1619454]
10. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*. 1987; 101:213–232.
11. Varni JW, Katz ER, Seid M, Quiggins DJL, FriedmanBender A, Castro CM. The pediatric cancer quality of life inventory (PCQL): I Instrument development, descriptive statistics, and cross-informant variance. *Journal of Behavioral Medicine*. 1998; 21:179–204. [PubMed: 9591169]
12. Upton P, Lawford J, Eiser C. Parent-child agreement across child health-related quality of life instruments: A review of the literature. *Quality of Life Research*. 2008; 17:895–913. [PubMed: 18521721]
13. Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children's health-related quality of life: An analysis of 13, 878 parents' reliability and validity across age subgroups using the PedsQL™ 4.0 Generic Core Scales. *Health and Quality of Life Outcomes*. 2007; 5(2):1–10. [PubMed: 17201920]
14. Campo JV, Comer DM, Jansen-McWilliams L, Gardner W, Kelleher KJ. Recurrent pain, emotional distress, and health service use in childhood. *Journal of Pediatrics*. 2002; 141:76–83. [PubMed: 12091855]
15. Janicke DM, Finney JW, Riley AW. Children's health care use: A prospective investigation of factors related to care-seeking. *Medical Care*. 2001; 39:990–1001. [PubMed: 11502956]
16. Varni JW, Setoguchi Y. Screening for behavioral and emotional problems in children and adolescents with congenital or acquired limb deficiencies. *American Journal of Diseases of Children*. 1992; 146:103–107. [PubMed: 1531284]
17. Varni JW, Burwinkle TM, Lane MM. Healthrelated quality of life measurement in pediatric clinical practice: An appraisal and precept for future research and application. *Health and Quality of Life Outcomes*. 2005; 3(34):1–9. [PubMed: 15634354]
18. Reise SP, Waller NG. Item response theory and clinical measurement. *Annual Review of Clinical Psychology*. 2009; 5:27–48.
19. Embretson, SE.; Reise, SP. *Item response theory for psychologists*. Erlbaum; Mahwah, NJ: 2000.
20. Irwin, DE.; Gross, HE.; Stucky, BD.; Thissen, D.; DeWitt, EM.; Lai, JS.; Amtmann, D.; Khastou, L.; Varni, JW.; DeWalt, DA. Development of the PROMIS® pediatrics proxy-report item banks. 2011. Manuscript under review
21. Muthén, LK.; Muthén, BO. *Mplus user's guide [Computer Software]*. 5th ed.. Muthén & Muthén; Los Angeles, CA: 2007.
22. Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*. 2010; 92:544–559. [PubMed: 20954056]
23. Cai, L.; du Toit, SHC.; Thissen, D. *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling [Computer software]*. Scientific Software International; Chicago, IL: inpress

24. Chen WH, Thissen D. Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*. 1997; 22:265–289.
25. Lord, FM. A study of item bias using item characteristic curve theory. In: Portinga, YH., editor. *Basic problems in crosscultural psychology*. Swets and Zeitlinge; Amsterdam: 1977. p. 19-29.
26. Cai L. SEM of another flavour: Two new applications of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology*. 2008; 61:309–329. [PubMed: 17971266]
27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*. 1995; 57:289–300.
28. Steinberg L, Thissen D. Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods*. 2006; 11:402–415. [PubMed: 17154754]
29. Orlando M, Thissen D. Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*. 2003; 27:289–298.
30. Thissen, D.; Nelson, L.; Rosa, K.; McLeod, LD. Item response theory for items scored in more than two categories. In: Thissen, D.; Wainer, H., editors. *Test scoring*. Lawrence Erlbaum Associates; Mahwah, NJ: 2001. p. 141-186.
31. Eiser C, Morse R. Can parents rate their child's health-related quality of life? Results from a systematic review. *Quality of Life Research*. 2001; 10:347–357. [PubMed: 11763247]

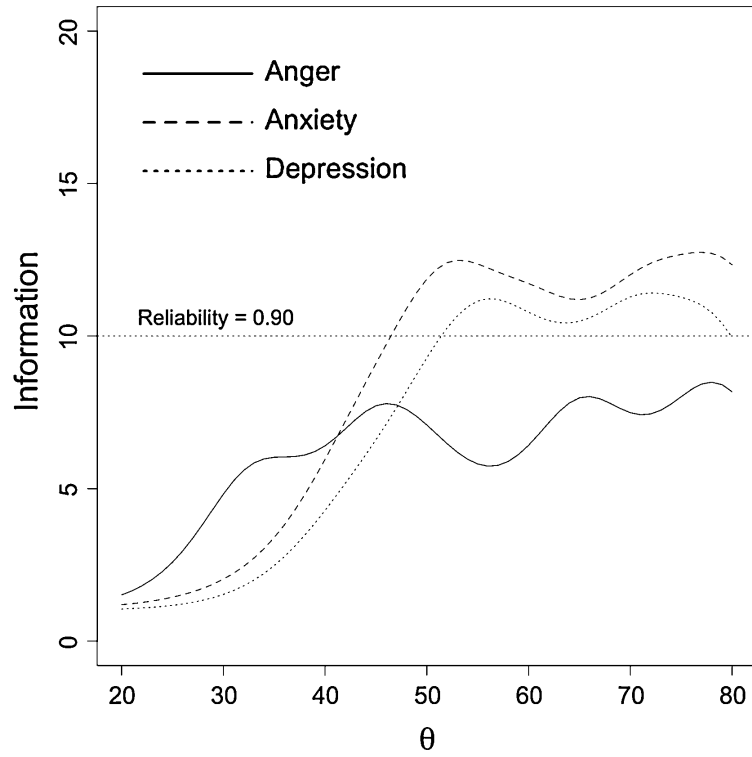


Fig. 1. Test information functions for the parent proxy PROMIS® domain short forms for the depressive symptoms, anxiety, and anger content domains

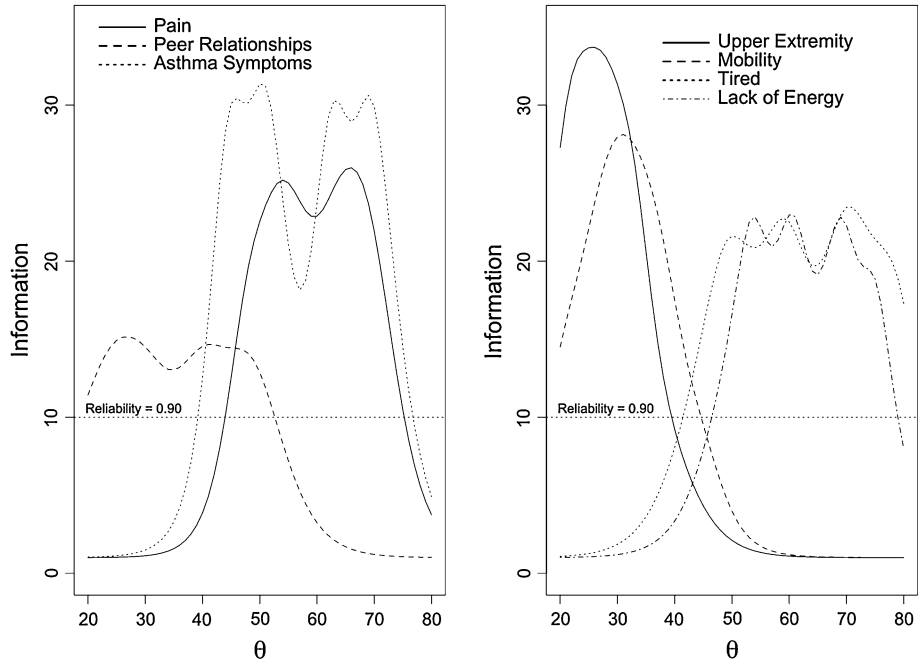


Fig. 2. Test information functions for the pain, peer relationships, asthma symptoms, upper extremity/dexterity, mobility, tiredness, and lack of energy content domains

Table 1

Parent and child demographics

	N = 1,548 (% of complete data)
Parent's gender	
Male	228 (15)
Female	1,313 (85)
Missing	7
Parent's age	
	Mean = 41.1, SD = 7.8
Marital status	
Never married	122 (8)
Married	1,060 (69)
Living with partner	67 (4)
Separated or divorced	256 (17)
Widowed	23 (2)
Missing	20
Parent's race	
White	980 (64)
Black or African-American	337 (22)
American Indian/Alaska Native	22(1)
Asian	30 (2)
Native Hawaiian/Pacific Is	5 (.3)
Other	107 (7)
Multiple races	50 (3)
Missing	17
Parent's ethnicity	
Non Hispanic	1,370 (89)
Hispanic	167 (11)
Missing	11
Guardian's relationship to child	
Mother, stepmother, foster mother	1,248 (81)
Father, stepfather, foster father	211 (14)
Grandparent	42 (3)
Guardian or other	35 (2)
Missing	12
Guardian's education level	
8th grade	27 (2)
Some high school	75 (5)
High school degree/GED	277 (18)
Some college/technical degree	529 (35)
College degree	433 (28)

	N = 1,548 (% of complete data)
Advanced degree	193 (13)
Missing	14
Child's age (years)	
8–12	665 (54)
13–17	566 (46)
Child's gender	
Male	736 (48)
Female	809 (52)
Missing	3

Table 2

Means and standard deviations of the pediatric self-report measures in the matched parent–child samples, and the correlations between latent variables measured by the pediatric self-report and parent proxy-report scales

Domain	Mean	Standard deviation	Parent–child correlation
Depressive symptoms (14 items)	48.9	10.1	0.37
Anxiety (15 items)	49.2	9.7	0.41
Anger (5 items)	50.1	10.2	0.40
Lack of energy (11 items)	49.9	9.8	0.41
Tired (23 items)	48.4	10.9	0.43
Upper extremity/dexterity (29 items)	49.3	11.6	0.69
Mobility (23 items)	50.4	12.2	0.63
Pain interference (13 items)	47.7	11.1	0.49
Peer relations (15 items)	49.8	10.5	0.35
Asthma impact (17 items)	46.9	11.9	0.53

Table 3

Explained common variance (ECV) for the general factor for each of the parent proxy scales, for the form combinations that were administered to parents

Scale	Form combination	ECV	Form combination	ECV
Depressive symptoms	1-2	0.89		
Anxiety	1-2	0.87		
Anger	1	1.0 [†]		
Lack of energy	3-4	0.82		
Tired	1-4	0.90	2-3	0.90
Upper extremity/dexterity	1-2	1.0 [†]	3-4	1.0 [†]
Mobility	1-2	0.96	3-4	1.0 [†]
Pain interference	1-4	0.93	2-3	0.98
Peer relations	1-2	0.89	3-4	0.87
Asthma impact	-	0.99		

[†]ECV values are 1.0 for unidimensional models

Table 4

Emotional distress item parameters

Short form (SF)	Item stem	Item parameters				S-X ² fit index			Sex DIF			Age DIF			
		a	b ₁	b ₂	b ₃	b ₄	χ^2	(df)	p	χ^2	(df)	p	χ^2	(df)	p
<i>Depressive symptoms</i>															
SF	My child felt unhappy ^c	3.23	-0.75	0.31	1.71	2.69	42(21)	0.00	7.8(5)	0.17	7.8(5)	0.17	7.8(5)	0.17	
SF	My child could not stop feeling sad	2.98	0.59	1.39	2.14	2.82	23(18)	0.17	6.0(5)	0.31	4.3(5)	0.50	4.3(5)	0.50	
SF	My child felt everything in his/her life went wrong ^b	2.75	0.26	0.91	1.89	2.65	36(28)	0.15	4.7(5)	0.46	2.1(5)	0.83	2.1(5)	0.83	
SF	It was hard for my child to do school work because he/she felt sad	2.70	0.65	1.46	2.28	2.91	20(19)	0.38	5.3(5)	0.38	4.5(5)	0.48	4.5(5)	0.48	
SF	It was hard for my child to have fun	2.52	0.07	1.04	2.15	2.92	31(24)	0.14	3.6(5)	0.60	6.1(5)	0.30	6.1(5)	0.30	
SF	My child felt sad ^c	2.43	-0.79	0.4	1.99	3.05	37(26)	0.07	4.2(5)	0.52	6.2(5)	0.28	6.2(5)	0.28	
SF	My child thought that his/her life was bad	2.26	0.06	0.94	2.19	3.13	38(29)	0.13	0.7(5)	0.98	5.1(5)	0.40	5.1(5)	0.40	
SF	My child felt alone ^a	2.23	0.40	1.38	2.76	3.16	26(21)	0.21	2.2(5)	0.82	4.2(5)	0.52	4.2(5)	0.52	
SF	My child felt like he/she could not do anything right ^b	2.18	-0.27	0.75	1.99	2.89	38(31)	0.19	-	-	4.0(5)	0.55	4.0(5)	0.55	
SF	My child felt lonely ^a	2.12	-0.12	0.87	2.28	3.82	43(30)	0.05	4.8(5)	0.44	8.7(5)	0.12	8.7(5)	0.12	
SF	My child felt too sad to eat	1.89	1.31	2.09	3.68	4.05	29(16)	0.03	-	-	-	-	-	-	
SF	My child felt stressed	1.84	-0.92	0.28	1.69	2.99	33(34)	0.52	4.0(5)	0.55	7.5(5)	0.18	7.5(5)	0.18	
SF	My child did not care about anything	1.14	0.58	1.88	3.54	4.82	37(35)	0.36	6.5(5)	0.27	8.8(5)	0.12	8.8(5)	0.12	
SF	My child wanted to be by himself/herself	0.89	-1.42	0.12	2.65	5.14	65(43)	0.02	2.8(5)	0.74	9.6(5)	0.09	9.6(5)	0.09	
<i>Anxiety</i>															
SF	My child felt scared	2.84	0.10	1.06	2.27	2.94	29(21)	0.11	5.3(5)	0.38	-	-	-	-	
SF	My child felt worried	2.65	-0.69	0.32	1.88	2.85	31(21)	0.11	1.7(5)	0.89	8.9(5)	0.11	8.9(5)	0.11	
SF	My child worried when he/she was at home ^e	2.63	0.28	1.08	2.53	3.26	24(22)	0.32	3.7(5)	0.59	-	-	-	-	
SF	My child worried when he/she went to bed at night ^e	2.51	0.28	1.01	2.12	3.11	39(24)	0.03	1.3(5)	0.94	1.6(5)	0.90	1.6(5)	0.90	
SF	My child felt like something	2.35	0.19	1.25	2.51	3.35	29(25)	0.25	4.3(5)	0.51	4.0(5)	0.55	4.0(5)	0.55	

Short form (SF)	Item stem	Item parameters				S-X ² fit index		Sex DIF		Age DIF		
		a	b ₁	b ₂	b ₃	b ₄	² (df)	p	² (df)	p	² (df)	p
	awful might happen ^f											
SF	My child worried about what could happen to him/her	2.24	-0.37	0.61	1.94	2.97	43(28)	0.04	4.6(5)	0.47	1.3(5)	0.93
SF	My child thought about scary things	2.19	-0.09	0.95	2.48	3.30	24(24)	0.45	3.0(5)	0.70	10.5(5)	0.06
	My child worried when he/she was away from home	1.96	0.55	1.57	2.55	2.84	32(23)	0.10	-	-	-	-
	My child got scared really easy	1.90	-0.05	1.23	2.46	3.35	45(28)	0.02	3.0(5)	0.70	6.6(5)	0.26
SF	My child felt nervous ^d	1.85	-0.82	0.34	1.92	3.21	29(31)	0.57	5.8(5)	0.33	3.1(5)	0.68
	My child woke up at night scared	1.80	0.71	1.82	3.04	3.33	36(22)	0.03	5.7(5)	0.34	-	-
	My child was worried he/she might die ^f	1.68	1.26	2.15	4.06		29(17)	0.04	8.8(5)	0.07	-	-
	It was hard for my child to relax ^d	1.58	-0.38	0.76	2.21	3.26	61(36)	0.01	8.9(5)	0.11	5.5(5)	0.35
	My child was afraid of going to school	1.53	1.23	2.21	3.18	3.56	22(21)	0.39	2.2(5)	0.82	8.8(5)	0.12
SF	My child was afraid that he/she would make mistakes ^d	1.27	-0.85	0.57	2.44	4.01	41(37)	0.28	6.2(5)	0.29	7.4(5)	0.19
	<i>Anger</i>											
SF	My child felt mad	3.16	-1.75	-0.51	1.53	2.85	13(13)	0.47	1.2(5)	0.95	6.9(5)	0.23
SF	My child was so angry he/she felt like yelling at somebody	2.42	-1.15	-0.05	1.55	2.72	40(21)	0.01	3.0(5)	0.70	5.8(5)	0.32
SF	My child was so angry he/she felt like throwing something	2.21	0.08	1.03	2.55	3.34	35(16)	0.00	6.3(5)	0.28	3.6(5)	0.61
SF	My child felt upset	2.02	-1.67	-0.33	2.02	3.34	19(20)	0.50	4.4(5)	0.50	6.5(5)	0.27
SF	When my child got mad, he/she stayed mad	1.77	-0.08	1.41	2.72	3.49	22(17)	0.19	4.9(5)	0.43	8.7(5)	0.12

^a indicate potentially locally dependent pairs of items. Items "My child felt alone" and "My child felt unhappy" were set aside from the final item pool for local dependence. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

^b indicate potentially locally dependent pairs of items. Items "My child felt alone" and "My child felt unhappy" were set aside from the final item pool for local dependence. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

- ^c indicate potentially locally dependent pairs of items. Items “My child felt alone” and “My child felt unhappy” were set aside from the final item pool for local dependence. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations
- ^d indicate potentially locally dependent pairs of items. Items “My child felt alone” and “My child felt unhappy” were set aside from the final item pool for local dependence. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations
- ^e indicate potentially locally dependent pairs of items. Items “My child felt alone” and “My child felt unhappy” were set aside from the final item pool for local dependence. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations
- ^f indicate potentially locally dependent pairs of items. Items “My child felt alone” and “My child felt unhappy” were set aside from the final item pool for local dependence. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

Table 5

Fatigue item parameters

Short form (SF)	Item stem	Item parameters				S- χ^2 fit index		Sex DIF		Age DIF		
		a	b ₁	b ₂	b ₃	b ₄	χ^2	(df)	p	χ^2	(df)	p
<i>Lack of energy</i>												
SF	My child had enough energy to do the things he/she likes to do	5.66	0.35	1.06	1.90	2.51	32(16)	0.01	23.8(5)	0.00	2.7(5)	0.74
SF	My child felt strong (not weak)	2.99	0.06	0.91	2.10	2.71	35(21)	0.03	8.2(5)	0.15	2.7(5)	0.74
SF	My child had enough energy to focus on his/her work ^a	2.87	-0.01	0.70	1.60	2.05	30(27)	0.29	3.4(5)	0.63	4.9(5)	0.43
	My child felt full of energy ^b	2.66	-0.32	0.62	1.45	2.10	52(30)	0.01	1.2(5)	0.94	5.8(5)	0.33
SF	My child had enough energy to do things outside	2.48	-0.02	0.57	1.31	1.73	39(26)	0.05	7.9(5)	0.16	7.2(5)	0.21
SF	My child had enough energy to go out with his/her family ^a	2.63	0.18	0.90	1.72	2.10	41(25)	0.03	1.0(5)	0.96	10.6(5)	0.06
SF	My child had enough energy to do sports or exercise	2.37	0.11	1.08	2.05	2.73	53(33)	0.02	3.5(5)	0.62	3.3(5)	0.65
SF	My child had enough energy to play or go out with his/her friends	2.39	0.19	0.85	1.58	1.97	34(28)	0.21	8.5(5)	0.13	1.9(5)	0.87
SF	My child had energy ^b	2.57	0.41	0.96	1.54	1.84	48(24)	0.00	3.2(5)	0.67	3.9(5)	0.57
	My child had enough energy to read	1.42	0.22	1.07	2.24	3.01	49(34)	0.04	4.1(5)	0.54	13.1(5)	0.02
	My child had enough energy to take a bath or shower ^a	1.24	0.99	1.74	2.44	2.79	76(26)	0.00	1.2(5)	0.95	7.0(5)	0.22
<i>Tired</i>												
SF	My child was too tired to enjoy the things he/she likes to do	4.36	-0.10	0.96	1.98	2.93	21(16)	0.17	-	-	-	-
SF	Being tired made it hard for my child to play or go out with friends as much as he/she would liked	3.43	0.21	1.01	2.23	2.72	17(17)	0.45	-	-	4.3(5)	0.50
SF	My child had trouble starting things because he/she was too tired	3.10	-0.38	0.72	1.87	2.59	38(27)	0.09	3.3(5)	0.66	1.8(5)	0.88
SF	My child felt weak ^b	2.75	0.06	0.87	1.96	3.05	33(26)	0.15	0.9(5)	0.97	-	-
	Being tired kept my child from having fun ^e	2.73	0.06	1.03	2.22	2.99	33(23)	0.09	1.7(5)	0.89	-	-

Short form (SF)	Item stem	Item parameters				S-X ² fit index				Sex DIF				Age DIF			
		a	b ₁	b ₂	b ₃	b ₄	² (df)	p	² (df)	p	² (df)	p	² (df)	p	² (df)	p	
SF	Being tired made it hard for my child to keep up with schoolwork	2.67	0.02	0.75	1.68	2.30	55(33)	0.01	2.2(5)	0.82	6.1(5)	0.30					
SF	My child got tired easily	2.61	-0.68	0.36	1.51	2.32	30(27)	0.3	1.4(5)	0.92	5.0(5)	0.42					
SF	My child was too tired to do sports or exercise ^e	2.48	-0.31	0.62	1.72	2.40	31(33)	0.58	3.0(5)	0.71	8.4(5)	0.13					
SF	My child was so tired it was hard for him/her to pay attention	2.43	-0.46	0.76	1.88	2.60	37(36)	0.42	3.9(5)	0.56	3.2(5)	0.67					
SF	My child had trouble finishing things because he/she was too tired	2.13	-0.54	0.66	1.82	2.40	39(37)	0.37	4.0(5)	0.55	10.1(5)	0.07					
	My child felt tired	2.13	-1.53	0.32	1.29	2.29	48(36)	0.09	3.3(5)	0.66	9.8(5)	0.08					
	My child was too tired to watch television ^c	2.03	0.65	1.79	2.98	-	27(20)	0.13	4.4(4)	0.35	4.7(4)	0.33					
	My child was too tired to focus on his/her work ^c	2.00	-0.64	0.64	2.12	2.93	34(29)	0.23	6.2(5)	0.29	6.9(5)	0.23					
	My child felt too tired to spend time with his/her friends ^e	1.97	0.33	1.39	2.58	3.12	32(25)	0.15	3.7(5)	0.60	6.3(5)	0.28					
	My child was too tired to go out with his/her family ^e	1.94	0.27	1.37	2.82	4.08	43(26)	0.02	-	-	-	-					
SF	My child was too tired to do things outside ^e	1.91	-0.18	0.89	2.13	2.82	26(33)	0.81	6.3(5)	0.28	5.0(5)	0.42					
	My child felt more tired than usual when he/she woke up in the morning	1.82	-0.81	0.44	1.66	2.64	46(42)	0.29	1.3(5)	0.93	1.5(5)	0.91					
	My child was too tired to go up and down a lot of stairs	1.81	0.31	1.07	2.06	2.51	40(30)	0.10	9.0(5)	0.11	9.6(5)	0.09					
	My child was too tired to eat	1.58	1.08	2.14	3.44	-	29(17)	0.04	5.8(4)	0.21	3.5(4)	0.48					
	My child needed to sleep during the day	1.50	-0.37	0.62	2.03	3.09	40(34)	0.22	8.0(5)	0.16	18.4(5)	0.00					
	My child was too tired to read	1.46	-0.12	0.98	2.48	3.45	38(32)	0.2	13.7(5)	0.02	10.0(5)	0.08					
	It was hard for my child to get out of bed in the morning because he/she was too tired	1.20	-0.65	0.53	2.24	3.18	56(39)	0.04	9.7(5)	0.08	11.5(5)	0.04					
	My child was too tired to take a bath or shower	1.15	0.58	1.97	4.15	5.73	28(25)	0.29	-	-	-	-					

^a indicate potentially locally dependent pairs or clusters of items. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. Dashes indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

- b* indicate potentially locally dependent pairs or clusters of items. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations
- c* indicate potentially locally dependent pairs or clusters of items. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations
- d* indicate potentially locally dependent pairs or clusters of items. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations
- e* indicate potentially locally dependent pairs or clusters of items. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

Table 6

Physical functioning item parameters

Short form (SF)	Item stem	Item parameters				S- χ^2 fit index		Sex DIF		Age DIF		
		a	b ₁	b ₂	b ₃	b ₄	χ^2 (df)	p	χ^2 (df)	p	χ^2 (df)	p
<i>Upper extremity</i>												
SF	My child could pull a shirt on over his/her head without help	4.83	-3.35	-2.89	-2.42	-2.05	-	-	-	-	6.0(5)	0.31
SF	My child could put on his/her shoes without help	4.83	-2.81	-2.52	-2.23	-1.82	7(3)	0.06	14.2(5)	0.01	6.6(5)	0.25
	My child could zip up his/her clothes	4.82	-3.04	-2.96	-2.43	-1.93	-	-	-	-	-	-
	My child could put toothpaste on his/her toothbrush without help	4.76	-3.38	-3.11	-2.70	-2.43	-	-	-	-	-	-
SF	My child could button his/her shirt or pants	4.64	-2.75	-2.63	-2.24	-1.66	8(6)	0.23	-	-	1.0(5)	0.97
	My child could put on his/her clothes without help	4.30	-3.10	-2.75	-2.35	-1.89	5(3)	0.16	11.3(5)	0.04	4.3(5)	0.51
	My child could pull on and fasten his/her seatbelt	4.19	-3.54	-3.29	-2.90	-2.12	12(2)	0.00	-	-	-	-
	My child could put on his/her socks without help	3.85	-2.99	-2.58	-2.14	-1.82	16(4)	0.00	14.9(5)	0.01	1.4(5)	0.93
	My child could cut paper with scissors	3.75	-3.00	-2.56	-1.88	-	7(6)	0.35	17.7(4)	0.00	7.9(4)	0.09
	My child could open his/her clothing drawers	3.73	-3.40	-3.23	-3.05	-2.35	-	-	-	-	-	-
	My child could hold a full cup	3.67	-3.86	-3.29	-3.03	-2.27	17(1)	0.00	45.6(5)	0.00	-	-
	My child could lift a cup to drink	3.53	-3.06	-2.67	-	-	-	-	-	-	3.5(3)	0.32
	My child could use a mouse or touch pad for the computer	3.48	-3.39	-2.81	-2.39	-	-	-	-	-	-	-
	My child could wash his/her face with a cloth	3.25	-3.01	-2.47	-	-	3(1)	0.07	-	-	2.4(3)	0.49
SF	My child could use a key to unlock a door	3.14	-2.90	-2.74	-2.50	-1.80	15(2)	0.00	28.3(5)	0.00	2.2(5)	0.82
SF	My child could open the rings in school binders	3.12	-3.37	-2.93	-2.27	-1.70	13(5)	0.02	37.8(5)	0.00	1.9(5)	0.87
SF	My child could pour a drink from a full pitcher	3.02	-2.84	-2.53	-2.08	-1.17	19(10)	0.04	11.7(5)	0.04	4.3(5)	0.51
	My child could tie shoelaces without help	2.80	-2.51	-2.30	-2.03	-1.61	18(10)	0.06	7.0(5)	0.23	5.5(5)	0.36

Short form (SF)	Item stem	Item parameters				S- χ^2 fit index		Sex DIF		Age DIF		
		a	b ₁	b ₂	b ₃	b ₄	χ^2	df	p	χ^2	df	p
	My child could dry his/her back with a towel	2.77	-3.03	-2.77	-2.20	-1.43	17(11)	0.10	15.0(5)	0.01	1.6(5)	0.90
	My child needed help with a bath	2.42	-2.95	-2.73	-2.22	-1.82	27(8)	0.00	22.4(5)	0.00	2.7(5)	0.74
	My child could turn door handles without help	2.41	-4.04	-3.51	-2.96	-2.19	11(3)	0.01	-	-	0.8(5)	0.98
SF	My child could open a jar by himself/herself	2.39	-2.83	-2.34	-1.72	-1.01	25(17)	0.10	19.7(5)	0.00	1.7(5)	0.89
SF	My child could pull open heavy doors	2.34	-2.81	-2.50	-2.01	-1.15	14(13)	0.38	33.1(5)	0.00	4.6(5)	0.47
	My child could dial a phone	2.30	-3.75	-3.58	-3.15	-2.79	-	-	-	-	-	-
	My child could hold an empty cup	2.16	-4.25	-3.25	-	-	-	-	-	-	-	-
	My child could move his/her hands or fingers	2.13	-3.45	-2.76	-2.18	-	10(7)	0.17	10.2(4)	0.04	10.7(4)	0.03
	My child could brush his/her teeth without help	2.07	-4.00	-3.52	-3.17	-2.55	5(1)	0.02	-	-	-	-
	My child could write with a pen or pencil	1.89	-4.04	-2.75	-2.10	-	13(6)	0.04	-	-	6.8(4)	0.15
	My child used a pencil with a special grip to write	0.92	-4.06	-3.67	-2.83	-2.37	23(12)	0.03	14.6(5)	0.01	3.0(5)	0.71
	<i>Mobility</i>											
	My child could get into bed by himself/herself	4.73	-3.38	-2.91	-2.28	-1.84	-	-	-	-	2.3(5)	0.81
	My child could walk across the room	4.52	-2.78	-2.57	-2.18	-1.67	9(1)	0.00	-	-	1.2(5)	0.95
	My child could bend over to pick something up	4.18	-3.08	-2.63	-2.03	-1.33	9(4)	0.06	17.8(5)	0.00	3.3(5)	0.65
	My child could get in and out of a car	4.13	-3.17	-2.86	-2.16	-1.50	14(4)	0.01	4.6(5)	0.46	5.2(5)	0.39
SF	My child could get up from the floor	4.00	-2.98	-2.35	-1.94	-1.38	15(6)	0.02	2.8(5)	0.73	4.6(5)	0.46
SF	My child could stand up without help	3.90	-2.73	-2.53	-2.2	-1.76	5(1)	0.02	1.1(5)	0.96	3.0(5)	0.70
SF	My child could do sports and exercise that other kids his/her age could do ^a	3.79	-2.01	-1.67	-1.20	-0.53	36(21)	0.02	4.2(5)	0.53	2.2(5)	0.82
SF	My child could walk up stairs without holding on to anything	3.76	-2.10	-1.87	-1.47	-1.07	21(17)	0.22	3.3(5)	0.66	4.1(5)	0.54
	My child could walk more than one block	3.32	-2.46	-2.09	-1.66	-1.01	13(8)	0.10	3.2(5)	0.67	8.9(5)	0.11
SF	My child could stand up on his/her tiptoes	3.18	-2.23	-2.09	-1.8	-1.40	20(15)	0.16	4.1(5)	0.54	4.5(5)	0.48

Short form (SF)	Item stem	Item parameters				S- χ^2 fit index		Sex DIF		Age DIF		
		a	b ₁	b ₂	b ₃	b ₄	χ^2	df	p	χ^2	df	p
	My child could get up from a regular toilet	3.15	-2.88	-2.80	-2.58	-2.03	19(7)	0.01	-	-	-	-
	My child could get down on his/her knees without holding on to something	3.08	-2.46	-2.24	-1.78	-1.25	29(21)	0.12	6.6(5)	0.25	7.5(5)	0.18
SF	My child could move his/her legs	3.06	-3.75	-2.78	-2.13	-1.64	14(4)	0.01	3.0(5)	0.70	2.1(5)	0.84
	My child used a wheelchair to get around	2.90	-2.86	-2.67	-2.50	-2.43	-	-	1.20(5)	0.94	9.2(5)	0.10
	My child could get out of bed by himself/herself	2.75	-3.21	-3.01	-2.55	-1.85	9(2)	0.01	-	-	-	-
	My child could ride a bike	2.67	-1.76	-1.66	-1.39	-0.94	29(22)	0.14	1.6(5)	0.90	3.1(5)	0.69
SF	My child could keep up when he/she played with other kids	2.58	-2.57	-2.18	-1.68	-0.93	23(10)	0.01	3.7(5)	0.60	7.0(5)	0.22
	My child could go up one step	2.55	-3.26	-3.01	-2.48	-2.01	15(9)	0.10	-	-	3.6(5)	0.61
	My child could carry his/her books in a backpack	2.42	-2.94	-2.43	-2.03	-1.29	27(11)	0.00	4.4(5)	0.49	9.1(5)	0.11
	My child could run a mile	2.34	-1.51	-0.99	-0.32	0.48	24(24)	0.44	2.6(5)	0.76	0.7(5)	0.98
SF	My child has been physically able to do the activities he/she enjoys most ^d	1.92	-2.95	-2.49	-1.66	-0.92	34(26)	0.13	5.0(5)	0.42	6.0(5)	0.31
	My child used a walker, cane or crutches to get around	1.88	-3.27	-3.19	-3.05	-2.83	-	-	-	-	-	-
	My child could turn his/her head all the way to the side	1.55	-5.08	-4.07	-3.06	-2.43	15(5)	0.01	-	-	-	-

^aIndicates potentially locally dependent pairs of items. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. Dashes indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

Table 7

Pain interference item parameters

Short form (SF)	Item stem	Item parameters				S- χ^2 fit index		Sex DIF		Age DIF		
		a	b ₁	b ₂	b ₃	b ₄	χ^2 (df)	P	χ^2 (df)	P	χ^2 (df)	P
SF	It was hard for my child to have fun when he/she had pain ^a	4.17	-0.21	0.36	1.32	1.72	21(20)	0.38	0.7(5)	0.98	3.8(5)	0.57
SF	It was hard for my child to pay attention when he/she had pain ^a	3.84	-0.19	0.49	1.46	1.91	28(21)	0.15	71.4(5)	0.00	3.9(5)	0.57
SF	My child had trouble doing schoolwork when he/she had pain	3.54	0.23	0.80	1.65	2.08	45(23)	0.00	6.1(5)	0.30	4.4(5)	0.49
	It was hard for my child to remember things when he/she had pain	3.37	0.35	1.02	2.03	2.70	23(25)	0.56	3.3(5)	0.65	-	-
SF	My child had trouble sleeping when he/she had pain	3.20	0.05	0.76	1.53	2.07	38(23)	0.03	1.9(5)	0.86	1.8(5)	0.88
SF	It was hard for my child to run when he/she had pain	3.02	-0.08	0.47	1.10	1.50	32(27)	0.24	7.3(5)	0.20	5.7(5)	0.34
SF	It was hard for my child to stay standing when he/she had pain ^e	2.90	0.25	0.69	1.43	1.87	26(30)	0.69	5.0(5)	0.42	1.4(5)	0.93
	It was hard for my child to get along with other people when he/she had pain	2.73	-0.01	0.76	1.65	2.15	41(20)	0.00	5.1(5)	0.41	3.6(5)	0.61
SF	It was hard for my child to walk one block when he/she had pain ^a	2.68	0.46	0.97	1.58	1.92	38(21)	0.01	3.6(5)	0.61	6.1(5)	0.29
SF	My child felt angry when he/she had pain	2.53	0.16	0.80	1.76	2.28	40(30)	0.11	2.7(5)	0.74	2.1(5)	0.83
	My child hurt a lot	2.45	-0.01	0.79	1.65	2.36	42(20)	0.00	1.0(5)	0.96	4.7(5)	0.46
	My child hurt all over his/her body	2.15	0.66	1.43	2.20	2.42	37(28)	0.12	-	-	-	-
	My child missed school when he/she had pain	1.70	0.77	1.30	2.35	2.57	19(15)	0.22	3.8(5)	0.58	2.7(5)	0.75

^aIndicate potentially locally dependent pairs of items. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

^eIndicate potentially locally dependent pairs of items. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

Table 8

Peer relations item parameters

Short form (SF)	Item stem	Item parameters					S-X ² fit index		Sex DIF		Age DIF	
		a	b ₁	b ₂	b ₃	b ₄	² (df)	P	² (df)	P	² (df)	P
SF	Other kids wanted to be with my child	3.53	-2.53	-2.03	-0.96	-0.15	18(16)	0.31	2.4(5)	0.79	6.3(5)	0.28
	My child felt good about his/her friendships	3.41	-2.97	-2.31	-1.21	-0.32	20(12)	0.06	8.7(5)	0.12	4.2(5)	0.52
	My child was able to have fun with his/her friends	2.64	-2.6	-2.29	-1.43	-0.60	14(17)	0.65	3.7(5)	0.59	3.9(5)	0.56
SF	Other kids wanted to talk to my child ^a	2.63	-3.27	-2.41	-1.36	-0.45	21(16)	0.16	3.8(5)	0.58	3.7(5)	0.59
SF	My child was good at making friends	2.56	-2.78	-2.14	-0.95	-0.16	30(19)	0.06	6.5(5)	0.26	6.1(5)	0.30
SF	My child was able to count on his/her friends ^b	2.54	-2.88	-2.07	-0.79	0.13	25(21)	0.25	1.9(5)	0.86	3.1(5)	0.68
SF	My child felt accepted by other kids his/her age	2.49	-2.32	-1.93	-1.14	-0.22	30(19)	0.05	5.2(5)	0.40	3.6(5)	0.61
SF	My child and his/her friends helped each other out	2.32	-2.99	-2.54	-1.04	-0.12	21(16)	0.19	4.2(5)	0.52	3.5(5)	0.63
SF	Other kids wanted to be my child's friend ^a	2.30	-2.72	-2.09	-1.05	-0.07	22(20)	0.32	5.0(5)	0.42	5.8(5)	0.33
	My child liked being around other kids his/her age	2.20	-2.95	-2.43	-1.54	-0.63	29(19)	0.06	3.5(5)	0.62	9.9(5)	0.08
	My child was able to talk about everything with his/her friends ^b	2.12	-2.80	-2.00	-0.75	0.07	19(22)	0.68	3.2(5)	0.67	3.1(5)	0.69
	My child was a good friend	2.04	-4.03	-3.34	-2.01	-0.78	13(12)	0.35	-	-	-	-
	My child spent time with his/her friends	1.68	-3.01	-2.16	-0.63	0.63	36(24)	0.05	3.6(5)	0.62	1.3(5)	0.94
	My child shared with other kids (food, games, pens, etc.)	1.52	-3.63	-3.14	-1.75	-0.55	22(20)	0.36	5.0(5)	0.42	0.9(5)	0.97
	My child played alone and kept to himself/herself	1.05	-3.50	-2.53	-0.58	0.95	34(31)	0.33	2.7(5)	0.74	9.5(5)	0.09

^aIndicate potentially locally dependent pairs of items. The item "My child was able to talk about everything with his/her friends" was set aside from the final item pool. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

^bIndicate potentially locally dependent pairs of items. The item "My child was able to talk about everything with his/her friends" was set aside from the final item pool. The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

Table 9

Asthma impact item parameters

Short form (SF)	Item stem	Item parameters				S-X ² fit index		Sex DIF		Age DIF		
		a	b ₁	b ₂	b ₃	b ₄	² (df)	P	² (df)	P	² (df)	P
SF	My child's asthma bothered him/her	5.32	-0.55	0.06	1.18	1.92	53(34)	0.02	3.5(5)	0.63	1.7(5)	0.89
SF	My child had trouble breathing because of his/her asthma	4.47	-0.58	0.13	1.29	1.96	44(37)	0.19	6.0(5)	0.30	3.0(5)	0.70
SF	My child felt wheezy because of his/her asthma	4.39	-0.41	0.28	1.35	2.00	51(37)	0.06	5.4(5)	0.37	1.6(5)	0.90
SF	It was hard for my child to take a deep breath because of asthma	3.83	-0.52	0.09	1.43	2.27	44(40)	0.29	6.6(5)	0.25	3.6(5)	0.61
	My child was bothered by the amount of time he/she spent wheezing	3.57	-0.28	0.33	1.48	2.01	53(44)	0.16	0.7(5)	0.99	3.1(5)	0.69
	My child had asthma attacks	3.39	0.07	0.66	1.65	2.16	83(39)	0.00	7.3(5)	0.20	4.3(5)	0.51
SF	My child's chest felt tight because of asthma	3.34	-0.56	-0.03	1.23	1.85	48(48)	0.47	5.3(5)	0.38	4.0(5)	0.55
	My child's body felt bad when he/she was out of breath	3.33	-0.63	0.02	1.12	1.86	52(50)	0.38	0.2(5)	1.00	1.8(5)	0.87
SF	My child had trouble sleeping at night because of asthma	3.15	-0.14	0.37	1.48	2.35	61(48)	0.09	4.5(5)	0.48	5.6(5)	0.35
	My child was bothered by asthma when he/she was with friends	2.99	-0.41	0.26	1.54	2.92	95(47)	0.00	-	-	-	-
	My child coughed because of his/her asthma	2.67	-0.91	-0.17	1.01	1.92	60(51)	0.18	4.9(5)	0.43	0.9(5)	0.97
	My child got tired easily because of his/her asthma	2.63	-0.64	0.01	1.29	2.33	61(59)	0.41	1.8(5)	0.87	1.3(5)	0.93
SF	It was hard for my child to play sports or exercise because of asthma	2.51	-0.48	0.28	1.42	2.28	69(58)	0.16	4.5(5)	0.48	2.2(5)	0.30
	My child had trouble walking because of asthma	2.35	0.32	1.01	2.12	2.74	60(44)	0.05	2.3(5)	0.80	5.5(5)	0.36
SF	My child felt scared that he/she might have trouble breathing because of asthma	2.12	-0.32	0.51	1.56	2.21	65(62)	0.36	5.7(5)	0.33	1.7(5)	0.89
	My child missed school because of asthma	1.96	0.46	1.04	1.96	2.65	74(46)	0.01	2.2(5)	0.83	4.8(5)	0.44
	It was hard for my child to play with pets because of asthma	1.55	0.10	0.71	1.53	2.05	67(64)	0.36	4.9(5)	0.43	6.1(5)	0.30

The item parameters are for the graded model as it is written by Thissen et al. [30], not including a scale factor of 1.7 used by some others. *Dashes* indicate a statistical test was not computed, because there was a difference between groups in the number of response categories with non-zero observations

Table 10
Scoring tables for the depressive symptoms, anxiety, anger, lack of energy, and tired parent proxy scales

Summed Score	Depressive symptoms		Anxiety		Anger		Lack of energy		Tired	
	EAP[/x]	SD[/x]	EAP[/x]	SD[/x]	EAP[/x]	SD[/x]	EAP[/x]	SD[/x]	EAP[/x]	SD[/x]
0	36	6	34	6	29	5	38	6	34	5
1	42	4	38	5	34	4	44	4	39	4
2	45	4	41	4	38	4	46	4	42	3
3	48	4	44	4	41	4	48	3	44	3
4	50	3	46	4	44	4	50	3	45	3
5	52	3	48	3	47	4	51	3	47	3
6	54	3	49	3	50	4	52	3	48	2
7	55	3	51	3	53	4	53	3	49	2
8	57	3	52	3	55	4	54	2	50	2
9	59	3	54	3	58	4	55	2	51	2
10	60	3	55	3	61	4	56	2	52	2
11	62	3	56	3	63	4	57	2	53	2
12	64	3	58	3	66	4	58	2	54	2
13	65	3	59	3	68	4	59	2	55	2
14	67	3	61	3	70	4	60	2	56	2
15	68	3	62	3	73	4	61	2	57	2
16	70	3	64	3	75	4	62	2	58	2
17	72	3	65	3	77	4	63	2	59	2
18	73	3	66	3	80	4	63	2	60	2
19	75	3	68	3	82	4	64	2	61	2
20	77	3	69	3	85	4	65	2	62	2
21	78	3	71	3			66	2	63	2
22	80	3	72	3			67	2	64	2
23	83	4	73	3			68	2	65	2
24	86	4	75	3			69	2	66	2
25			76	3			70	2	67	2
26			77	3			71	2	68	2

Summed Score	Depressive symptoms		Anxiety		Anger		Lack of energy		Tired	
	EAP[<i>x</i>]	SD[<i>x</i>]	EAP[<i>x</i>]	SD[<i>x</i>]	EAP[<i>x</i>]	SD[<i>x</i>]	EAP[<i>x</i>]	SD[<i>x</i>]	EAP[<i>x</i>]	SD[<i>x</i>]
27			79	3			72	3	69	2
28			80	3			73	3	70	2
29			82	3			75	3	71	2
30			84	3			76	3	72	2
31			86	4			78	3	72	2
32			88	4			81	4	73	2
33									74	2
34									75	2
35									76	2
36									77	2
37									79	3
38									80	3
39									82	3
40									85	4

EAP expected a posteriori, SD standard deviation, latent construct, x summed score

Table 11

Scoring tables for the upper extremity/dexterity, mobility, pain interference, peer relations, and asthma impact parent proxy scales

Score	Upper extremity/dexterity		Mobility		Pain interference		Peer relations		Asthma impact	
	EAP[\bar{x}]	SD[\bar{x}]	EAP[\bar{x}]	SD[\bar{x}]	EAP[\bar{x}]	SD[\bar{x}]	EAP[\bar{x}]	SD[\bar{x}]	EAP[\bar{x}]	SD[\bar{x}]
0	13	3	14	4	38	6	15	4	32	6
1	16	3	17	3	44	3	18	3	39	4
2	17	3	20	3	46	3	20	3	41	3
3	18	2	21	3	48	3	22	3	43	3
4	19	2	22	3	49	2	23	3	44	2
5	20	2	23	2	50	2	24	3	46	2
6	21	2	24	2	51	2	26	3	47	2
7	22	2	25	2	52	2	27	3	48	2
8	22	2	26	2	53	2	28	3	49	2
9	23	2	27	2	54	2	29	3	50	2
10	24	2	27	2	55	2	31	3	51	2
11	24	2	28	2	56	2	32	3	52	2
12	25	2	29	2	57	2	33	3	53	2
13	25	2	29	2	58	2	34	3	54	2
14	26	2	30	2	58	2	36	3	55	2
15	26	2	31	2	59	2	37	3	56	2
16	27	2	31	2	60	2	38	3	58	2
17	28	2	32	2	61	2	39	3	59	2
18	28	2	33	2	62	2	41	3	60	2
19	29	2	33	2	62	2	42	3	61	2
20	30	2	34	2	63	2	43	3	63	2
21	30	2	35	2	64	2	45	3	64	2
22	31	2	35	2	65	2	46	3	65	2
23	32	2	36	2	66	2	48	3	66	2
24	33	2	37	2	67	2	49	3	67	2
25	34	3	38	3	67	2	51	3	68	2
26	35	3	39	3	68	2	53	4	69	2

Summed Score	Upper extremity/dexterity		Mobility		Pain interference		Peer relations		Asthma impact	
	EAP[λ]	SD[λ]	EAP[λ]	SD[λ]	EAP[λ]	SD[λ]	EAP[λ]	SD[λ]	EAP[λ]	SD[λ]
27	37	3	40	3	69	2	56	4	70	2
28	38	4	42	4	70	2	62	6	71	2
29	40	4	43	4	71	3			73	2
30	42	4	45	4	73	3			74	3
31	45	5	48	4	74	3			76	3
32	55	8	56	7	78	4			80	5

EAP expected a posteriori, SD standard deviation, latent construct, x summed score