PROMIS® Parent Proxy Report Scales for Children Ages 5–7 Years: An Item Response Theory Analysis of Differential Item Functioning across Age Groups

James W. Varni[1], **David Thissen**[2], **Brian D. Stucky**[3], **Yang Liu**[2], **Brooke Magnus**[2], **Hally Quinn**[2], **Debra E. Irwin**[4], **Esi Morgan DeWitt**[5], **Jin-Shei Lai**[6], **Dagmar Amtmann**[7], **Heather E. Gross**[8], and **Darren A. DeWalt**[8]

[1]Department of Pediatrics, College of Medicine, Department of Landscape Architecture and Urban Planning, College of Architecture, Texas A&M University, College Station, TX, USA

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

[3]RAND Corporation, Santa Monica, CA, USA

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

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

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

[7]Department of Rehabilitation Medicine, University of Washington, Seattle, WA, USA

[8]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, USA

Abstract

Objective—The objective of the present study is to describe the extension of the National Institutes of Health (NIH) Patient Reported Outcomes Measurement Information System (PROMIS®) pediatric parent proxy-report item banks for parents of children ages 5–7 years, and to investigate differential item functioning (DIF) between the data obtained from parents of 5–7 year old children with the data obtained from parents of 8–17 year old children in the original construction of the scales.

Methods—Item response theory (IRT) analyses of DIF were conducted comparing data from the 5–7 age group with data from the established scales for ages 8–17 across 5 generic health domains (physical functioning, pain, fatigue, emotional health, social health) and asthma.

Results—IRT DIF analyses revealed that the majority of the items functioned similarly with responses from parents of younger and older children. A small number of items were removed from the item bank for younger children, and a few items that exhibited statistical DIF were retained in the pools with the caveat that they should not be used in studies that involve comparisons of younger children with older children.

Address correspondence to: James W. Varni, Ph.D., Professor of Architecture and Medicine, College of Architecture, Texas A&M University, 3137 TAMU, College Station, Texas 77843-3137 USA. jvarni@arch.tamu.edu. Telephone: (979) 862-1095. Fax: (979) 862-1784.

Conclusions—The study confirms that most of the items in the PROMIS parent proxy-report item banks can be used with parents of children ages 5–7. It is anticipated that these new scales will have application for younger pediatric populations when 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 (PROs) in chronic diseases. Items are evaluated using Item Response Theory (IRT) to derive scales with scores that are theoretically 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 9 years, the PROMIS Pediatric Cooperative Group has developed pediatric self-report item banks for ages 8–17 years across five generic health domains (physical functioning, pain, fatigue, emotional health, 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, so generic or nondisease-specific scales were developed [4–10]. An asthma-specific measure was also created [11; 12].

While pediatric self-report should be considered the standard for measuring PROs [13], there may be circumstances when the child is too young, too cognitively impaired, or too ill to complete a PRO instrument, and parent proxy-report may be needed in such cases [14]. To address this need, we developed the PROMIS Parent Proxy Report Scales for Children [15]. Our initial report focused on children ages 8–17 years [16].

The majority of parent proxy-report scales, consistent with other clinical assessment instruments [17], have utilized Classical Test Theory (CTT) and have rarely taken advantage of IRT analysis in the scale development process [18]. 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 user can then select the number of items to administer and obtain scores on the same metric as all other users of this item bank [18].

Since our initial report on the PROMIS Parent Proxy Report Scales focused on children ages 8–17 years [16], the objective of the present study is to describe the extension of these parent proxy-report scales for parents of children ages 5–7 years. Because the PROMIS parent proxy-report item banks already exist for use with parents of children ages 8–17 [16], we do this by using the analysis of differential item functioning (DIF) between the data obtained from parents of 5–7 year old children with the data obtained from parents of 8–17 year old children in the original construction of the scales.

DIF analysis is a procedure designed to investigate whether items measure the same unobserved constructs in the same way in two groups. Group means on the construct may differ, but that may be due to group differences on the individual differences variable measured by the scale. IRT analysis makes use of parameters associated with each item; if the item parameter estimates differ significantly between groups, there is evidence that the items are not measuring the latent variable in the same way across groups; that is DIF.

The primary goal of the research reported here is to investigate DIF between responses from parents of children ages 5–7 years and those from parents of the 8–17 year old children of

the original sample. For items that do not exhibit DIF between responses from parents of children in the 5–7 age group and the original sample of data with children 8–17, we conclude that those items may be used in the extension of the scales to ages 5–7. Items that exhibit DIF may not measure the same constructs in the same ways with responses from parents of younger children as they do when the children are older; those items may be excluded from use with parents of children 5–7 years of age. In practice there are always shades of gray: Some items may exhibit statistical DIF, but often their practical effect on the overall score is negligible and if that's the case, they may be useful for measuring health outcomes for younger children. We examine parent proxy-report items for the presence of DIF across three age groups: 5–7, 8–12, and 13–17.

We examine DIF between parent proxy-report responses for the 5–7 and 8–17 age groups as the primary analysis, and between the 8–12 and 13–17 age groups in a secondary analysis. The primary analysis addresses the study's main research question. The purpose of the secondary analysis is twofold: First, it checks for DIF within the current 8–17 age range for the PROMIS parent proxy-report scales; if such DIF is present, it could complicate interpretation of the results of the primary DIF analysis that has the 8–17 age group aggregated. Second, if we find the expected absence of DIF between responses from parents of 8–12 and 13–17 year old children, but we do find DIF between the 5–7 and 8–17 groups, then we are reassured that the DIF reflects meaningful differences in performance of the items between parents of younger children and those of 8–17 year old children.

Methods

Participants

Participants were recruited between May 2008 through March 2009 in hospital-based outpatient general pediatrics and subspecialty clinics. Parent of pediatric patients within the age range of 5–17 were recruited through a review of clinic appointment rosters or while waiting for their clinic appointments according to protocols approved by the institutional review boards (IRBs) of University of North Carolina (UNC), Duke University Medical Center, University of Washington (UW), Lurie Children's Hospital of Chicago (Lurie; formally Children's Memorial Hospital, Chicago), and Children's Hospital at Scott and White (S&W) in Texas. Trained research assistants, who were under the supervision of one of the authors (Site Principal Investigator) and were based at the clinics, were responsible for reviewing the clinic appointment rosters. The trained research assistants were responsible for approaching potential participants and briefly explaining the study in clinic and hospital waiting rooms. The UNC, Duke, UW, Lurie 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. Parents of 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. Parents signed an informed consent document. Each participant received a $10 gift card in return for their time and effort.

Item Bank Development

The PROMIS Pediatric item banks were developed using a strategic item generation methodology adapted 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. Item development has been described in detail previously [15; 16].

The final pediatric self-report item banks included scales measuring five generic health domains (Physical Functioning, Pain, Fatigue, Emotional Health, Social Health) and Asthma. Because Physical Functioning includes both upper extremity and mobility item banks, Emotional Health includes separate anger, anxiety and depressive symptoms item banks, and Fatigue includes both tired and lack of energy item banks, a total of 10 content domains were tested [4–10].

The parent proxy-report items were developed from the 10 existing pediatric self-report content domains [4–10]. The items were revised to retain their meaning, while modifying the phrasing so that all items involved parents reporting on their child for the same 10 domains as the pediatric self-report domains [15; 16]. 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."

All items had a 7-day recall period and used one of two sets of standardized 5-point response options: *never*, *almost never*, *sometimes*, *often*, *almost always* for all scales except physical functioning; or, *with no trouble*, *with a little trouble*, *with some trouble*, *with a lot of trouble*, *not able to do* for the physical functioning scales.

In the data collection for the original standardization of the pediatric parent-proxy scales, 293 proxy-report items from the 10 content domains were administered to 432 parents of 5–7 year old children and 1,548 parents of the 8–17 year old children [15]. 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; the details of the sampling design have been described previously [15]. For children ages 5–7, all responses were provided by parent-proxy, whereas for ages 8–17, responses were provided both by parent-proxy and directly by the children (though only parent-proxy responses are use in the analyses presented here).

Of the 293 items administered, 165 were ultimately included in the proxy item banks for parents of children ages 8–17 years; these corresponded to the 166 items that were ultimately included in the pediatric self-report item banks, less one item that could not be re-worded for parent-proxy report. This strategy was taken to maximize the comparability between the pediatric self-report and parent proxy-report versions. The general process of reducing the pediatric self-report items from 293 to 166 has been previously reported [4]. 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 [16].

Statistical and Psychometric Methods

Traditional descriptive statistics were computed as a check on data entry and validity and to verify that there were no empty (zero frequency) response categories for any item. Items were analyzed for DIF after dividing the respondents into three age groups: ages 5–7, 8–12, and 13–17. Two DIF comparisons were made: ages 5–7 vs. ages 8–17, and ages 8–12 vs. ages 13–17. The DIF analyses were performed using the Wald Test as implemented in IRTPRO [19]. Significant χ^2 values indicate that the item parameters are different across the two groups (i.e., the items are not measuring the same construct or are functioning differently across groups). DIF statistics were computed only for parameters associated with

response categories that were endorsed across all ages. Items associated with slope parameter estimates exceeding 10 were omitted from the DIF analyses, because this also indicates insufficient data to obtain stable parameter estimates.

The Benjamini-Hochberg procedure [20] was used to adjust alpha levels for multiple comparisons. Standard methods were employed to determine the effect size of the DIF [21]; an item can show statistically significant DIF but still have a relatively small effect size, making its practical significance trivial. For this reason, it is important to examine exactly how the DIF affects responses across groups.

Finally, items exhibiting DIF were reviewed by an expert panel which consisted of 7 individuals with a range of expertise in the statistical techniques used, domain content, and the use of patient-reported outcomes in pediatric populations. The panel then recommended by consensus whether each of the DIF items would be removed from the scale for parent-proxy responses for children ages 5–7, or retained with a warning that the items should not be used in studies comparing 5–7 year old children with older children. Specifically, the expert panel was asked to both look at the results of the DIF analysis and consider the content of the item. If DIF was found, panel members considered the developmental appropriateness of the item for children ages 5–7. Each panel member decided whether the item exhibiting DIF was developmentally appropriate for children ages 5–7, and voted on whether to retain the item or remove it. Once all of the votes were tallied, the group discussed all items where the vote was not unanimous to reach a final consensus decision.

Results

The nine test forms were completed by a total of 1980 respondents (432 in the 5–7 age group and 1,548 in the 8–17 age group). Demographic information for the respondents can be found in Table 1. The total sample was 53% female and 22% of the children were ages 5–7. The caregivers providing responses from their children were 64% Caucasian, 22% Black, 3% multi-racial, and 11% other races (Asian/Pacific Islanders, Native Americans, and Other Races). Eleven percent of the sample was of Hispanic ethnicity.

Due to imbalance in the cross-site assignment of forms to respondents, the form originally numbered 4 (Form 4) was administered to too few respondents for IRT analysis; as a result, the subset of items derived from that form are not included in the present analysis.

Emotional Health

The Emotional Health domain consists of three subdomains: Depressive Symptoms, Anxiety, and Anger. Within the Depressive Symptoms domain, four items exhibit statistically significant DIF when comparing children ages 5–7 to those 8–17: "my child felt sad" (χ^2 (5) = 19.5, $p < 0.01$); "my child thought that his/her life was bad" (χ^2 (5) = 21.7, $p < 0.01$); "it was hard for my child to do school work because he/she felt sad" (χ^2 (5) = 16.2, $p = 0.01$); and "my child felt stressed" (χ^2 (5) = 19.5, $p < 0.01$); see Table 2. Of these four items, the first two are also on the parent proxy-report short form [16]. For the item "my child felt sad", computation of IRT-modeled response probabilities suggested only a weak effect size of the DIF. For the item "my child thought that his/her life was bad", DIF increased the number of responses at both ends of the scale. Parents of younger children were more likely to endorse the extreme response categories on this item, making it more discriminating for the younger children. After expert panel discussions, it was decided that the two short form items ("my child felt sad" and "my child thought that his/her life was bad") be retained for the Depressive Symptoms parent proxy-report scale for ages 5–7, but that they be flagged to exclude their use in comparison with children in older age groups. The other two items exhibiting DIF were omitted from the scale for ages 5–7, as these items

differ in their relevance depending on the age group. None of the Depressive Symptoms items showed DIF when comparing ages 8–12 vs. 13–17.

Five of the Anxiety items exhibit DIF when comparing ages 5–7 vs. 8–17: "my child worried when he/she was at home" (χ^2 (4) = 13.1, p = 0.01); "my child worried when he/she went to bed at night" (χ^2 (5) = 19.2, $p < 0.01$); "my child thought about scary things" (χ^2 (5) = 14.2, p = 0.01); "my child got scared really easy" (χ^2 (5) = 34.5, $p < 0.01$); and "my child woke up at night scared" (χ^2 (4) = 19.4, $p < 0.01$); see Table 2. Two of these items are also on the short form ("my child worried when he/she went to bed at night" and "my child thought about scary things"). The expert panel recommended omission of "my child got scared really easy" from the parent proxy-report scale for ages 5–7 due to its large effect size. Younger children may define "scared" differently from older children, and being scared is rather common in younger children, regardless of their health status. A similar rationale was used for the item "my child woke up at night scared"; therefore, this item was also omitted. The remaining three items showing statistical DIF were retained for the scale due to their weak effect sizes and their presence on the short form. Once again, none of the items exhibit DIF in the comparison of parent proxy-report responses with children ages 8–12 vs. 13–17.

The Anger item bank consists of six items; none of these showed statistically significant DIF for either age group comparison after controlling for multiplicity. All six items were retained for inclusion on the scale for all age groups.

Fatigue

The Fatigue domain comprises two subdomains: Lack of Energy and Tired. Four of the items on the Lack of Energy scale are on Form 4, and therefore, not included in the DIF analyses. See Table 3 for a list of these items. One other item ("my child had enough energy to do the things he/she likes to do") is not included in the analysis due to producing a slope parameter estimate larger than 10. Of the items included in the analysis, two Lack of Energy items show statistical DIF: "my child felt full of energy" (χ^2 (4) = 16.1 $p < 0.01$) and "my child had enough energy to go out or play with his/her friends" (χ^2 (5) = 20.9, $p < 0.01$). The expert panel recommended that both items be retained for the Lack of Energy scale but that they not be used in studies comparing children ages 5–7 with older children. In comparing ages 8–12 vs. 13–17, none of the items show statistical DIF.

Several Tired items are not included in the analysis because they are on Form 4; see Table 3 for a list of the nine items falling into this category. Of the items not on Form 4, four of them reveal statistically significant DIF: "my child was too tired to do sports or exercise" (χ^2 (5) = 17.6 $p < 0.01$); "my child felt too tired to spend time with his/her friends" (χ^2 (4) = 15.1 $p < 0.01$), "my child was too tired to eat" (χ^2 (4) = 12.7 p = 0.01), and "my child needed to sleep during the day" (χ^2 (5) = 16.8 $p < 0.01$); see Table 3. Of the items showing statistical DIF, only one ("my child needed to sleep during the day") was excluded from the scale for ages 5–7. This item is not very discriminating for 5–7 year-olds, likely due to some children still taking naps during the day regardless of their levels of fatigue. The other three items showing DIF were recommended for inclusion in the scale but not for administration in studies comparing 5–7 year-olds with older children. None of the items comparing ages 8–12 vs. 13–17 exhibited statistical DIF.

Physical Functioning

The Physical Functioning domain consists of two subdomains: Upper Extremity and Mobility. Thirteen Upper Extremity items are not included in the DIF analyses: seven from Form 4 due to insufficient sample size and six because of slope estimates larger than 10; see

Table 4 for a list of these items. Of the rest of the items on the scale, one item, "my child could dial a phone," exhibits statistical DIF when comparing children ages 5–7 to those 8–17 (χ^2 (4) = 17.0, $p < 0.01$). The expert panel decided that the item should be excluded from the scale for ages 5–7 because dialing a phone is not a usual activity for children this young. None of the items demonstrate statistical DIF for the 8–12 vs. 13–17 age comparison.

Thirteen Mobility items are not included in the DIF analyses: seven from Form 4 as a consequence of insufficient sample size, five due to slope estimates larger than 10, and one because participants responded in only one response category; see Table 4 for a list of these items. Of the items included in the analysis, three Mobility items show significant DIF when comparing children ages 5–7 to those 8–17: "my child could do sports and exercise that other kids his/her age could do" (χ^2 (5) = 17.3, $p < 0.01$), "my child could go up one step" (χ^2 (3) = 12.7, $p = 0.01$), and "my child could run a mile" (χ^2 (5) = 52.9, $p < 0.01$). In addition, the item "my child could get up from a regular toilet" (χ^2 (4) = 36.7, $p < 0.01$) exhibited statistical DIF when comparing children 8–12 vs. 13–17. The expert panel decided to retain all items showing statistical DIF except "my child could run a mile" for ages 5–7 because a mile is too long of a distance for this age group.

Pain Interference

Five items on the Pain Interference scale are on Form 4 so are excluded from the DIF analyses due to insufficient sample size; see Table 5 for a list of the items falling into this category. Of the items included in the analyses, none show DIF for either of the age group comparisons. All items are retained for all age groups.

Peer Relationships

Five items on the Peer Relationships scale are on Form 4 and therefore not included in the DIF analyses due to insufficient sample size; see Table 6 for a list of these items. Of the items included in the analysis, "other kids wanted to be my child's friend" is the only item to show statistical DIF when comparing children ages 5–7 to those 8–17 (χ^2 (3) = 30.1, $p < 0.01$). The content experts chose to retain the item but suggested that the item not be used in studies comparing 5–7 year olds to 8–17 year olds. The concept of friendship may be different in younger children than older children. None of the items exhibit statistical DIF when comparing ages 8–12 vs. 13–17.

Asthma

One Asthma item, "my child coughed because of his/her asthma," exhibited statistical DIF between 5–7 and 13–17 children (χ^2 (5) = 19.6, $p < 0.01$); see Table 7. The expert panel decided to retain this item despite the large effect size. It is unclear as to why this item demonstrates DIF but the item content of coughing does not seem to have different meanings based on a child's age. None of the items demonstrate DIF for the ages 8–12 vs. 13–17 comparisons.

Discussion

This study describes the extension of the NIH PROMIS Parent Proxy Report Scales to ages 5–7 based on IRT DIF analyses that suggested that the majority of the items functioned similarly when responses from parents of younger children were compared with those from the original sample of parents of children ages 8–17 years. A small number of items that are useful with parents of children ages 8–17 years were removed from the item bank for younger children, due to the results of the DIF analysis, but largely because they involve terms (like "stress") or activities (like "homework") that have different meaning, or no meaning, for younger children. A few items that exhibited statistical DIF were retained in

the item pools for administration to parents of children ages 5–7, but with the caveat that they should not be used in studies that involve comparisons of younger children with older children, because these items function differently for younger children than for older children, although the items are effective indicators at all ages.

Children aged 5–7 have very different life experiences from older children, not only developmental differences but also socio-environmental differences (e.g., school experiences). Thus, we expected parents of younger children might perceive some items differently from those of older children. Understanding the similarities and differences between parents' perceptions for these age groups is important. Items that showed measurement equivalence between age groups can be administered across the age groups, while items that demonstrated DIF should only be used to capture the developmental uniqueness of each age group.

A secondary DIF analyses also compared the performance of items for children ages 8–12 with those 13–17. Virtually no DIF was observed in these comparisons across the age range already covered by the pre-existing pediatric self-report and parent proxy-report scales. That is a reassuring result for the validity of the existing scales, and supports our use of the 8–17 year old group to serve as a monolithic comparison group for the 5–7 year old group in the primary DIF analysis. The contrast between the DIF found in the primary and secondary analysis also supports the degree to which we attend to DIF findings when parent proxy-reports for younger children are considered.

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, so we cannot assume that the scales would have the same test characteristics in those other populations. Further, we were not able to conduct IRT analyses with the Form 4 items due to insufficient sample size. Finally, disease diagnoses were not verified from the medical charts, but rather were based on parent report. Future research will include patients with verified disease diagnoses.

Future research with other samples may identify 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 and extends the NIH PROMIS Parent Proxy Report Scales for ages 5–7. 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

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, 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). Medical Care. 2007; 45 (Suppl 1):S22–S31. [PubMed: 17443115]

3. 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. Medical Care. 2007; 45 (Suppl 1):S3–S11. [PubMed: 17443116]

4. Irwin DE, Stucky BD, Thissen D, DeWitt EM, Lai JS, Yeatts K, Varni JW, DeWalt DA. 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, Varni JW, Yeatts K, DeWalt DA. 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, Yeatts K, DeWalt DA. 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; 64:794–804. [PubMed: 21292444]

8. Irwin DE, Stucky BD, Langer MM, Thissen D, DeWitt EM, Lai JS, Yeatts KB, Varni JW, DeWalt DA. PROMIS Pediatric Anger Scale: An item response theory analysis. Quality of Life Research. 2012; 21:697–706. [PubMed: 21785833]

9. DeWalt DA, Thissen D, Stucky BD, Langer MM, DeWitt EM, Irwin DE, Lai JS, Yeatts KB, Gross HE, Taylor O, Varni JW. PROMIS Pediatric Peer Relationships Scale: Development of a peer relationships item bank as part of social health measurement. Health Psychology. in press.

10. Lai JS, Stucky BD, Thissen D, Varni JW, DeWitt EM, Irwin DE, Yeatts KB, Dewalt DA. Development and psychometric properties of the PROMIS® pediatric fatigue item banks. Quality of Life Research. in press.

11. Yeatts K, Stucky BD, Thissen D, Irwin DE, Varni JW, DeWitt EM, Lai JS, DeWalt DA. 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]

12. Thissen D, Varni JW, Stucky BD, Liu Y, Irwin DE, DeWalt DA. Using the PedsQL 3.0 Asthma Module to obtain scores comparable with those of the PROMIS Pediatric Asthma Impact Scale (PAIS). Quality of Life Research. 2011; 20:1497–1505. [PubMed: 21384264]

13. Varni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life?: An analysis of 8,591 children across age subgroups with the PedsQL 4.0 Generic Core Scales. Health and Quality of Life Outcomes. 2007; 5:1, 1–13. [PubMed: 17201920]

14. 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: 17201923]

15. Irwin DE, Gross HE, Stucky BD, Thissen D, DeWitt EM, Lai JS, Amtmann D, Khastou L, Varni JW, DeWalt DA. Development of six PROMIS pediatrics parent proxy-report item banks. Health and Quality of Life Outcomes. 2012; 10:22, 1–13. [PubMed: 22357192]

16. Varni JW, Thissen D, Stucky BD, Liu Y, Gorder H, Irwin DE, DeWitt EM, Lai JS, Amtmann D, DeWalt DA. PROMIS® Parent Proxy Report Scales: An Item Response Theory analysis of the parent proxy report item banks. Quality of Life Research. 2012; 21:1223–1240. [PubMed: 21971875]

17. Reise SP, Waller NG. Item response theory and clinical measurement. Annual Review of Clinical Psychology. 2009; 5:27–48.

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

19. Cai, L.; du Toit, SHC.; Thissen, D. IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling [Computer software]. Chicago, IL: Scientific Software International; in press

20. 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.

21. 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]

Table 1

Parent and child demographics

	Ages 5–7 N = 432 (% of complete data)	Ages 8–17 N = 1,548 (% of complete data)
Caregiver's Gender		
Male	66 (15)	228 (15)
Female	365 (85)	1,313 (85)
Missing	1	7
Caregiver's Age	Mean = 37.5, SD = 7.9	Mean = 41.1, SD = 7.8
Marital Status		
Never married	49 (11)	122 (8)
Married	306 (71)	1,060 (69)
Living with partner	23 (5)	67 (4)
Separated or divorced	46 (11)	46 (11)
Widowed	5 (1)	5 (1)
Missing	3	20
Caregiver's Race		
White	271 (64)	980 (64)
Black or African-American	94 (22)	337 (22)
American Indian/Alaska Native	3 (1)	22 (1)
Asian	10 (2)	30 (2)
Native Hawaiian/Pacific Is	1 (.2)	5 (.3)
Other	34 (8)	107 (7)
Multiple Races	12 (3)	50 (3)
Missing	7	17
Caregiver's Ethnicity		
Non Hispanic	381 (89)	1,370 (89)
Hispanic	47 (11)	167 (11)
Missing	4	11
Caregiver's Relationship to Child		
Mother, stepmother, foster mother	352 (82)	1,248 (81)
Father, stepfather, foster father	61 (14)	211 (14)
Grandparent	11 (3)	42 (3)
Guardian or other	2 (.5)	35 (2)
Missing	4	12
Caregiver's Education Level		
<= 8th grade	3 (1)	27 (2)
Some high school	23 (5)	75 (5)
High school degree/ GED®	71 (17)	277 (18)
Some college/technical degree	126 (29)	529 (35)
College degree	123 (29)	433 (28)
Advanced degree	82 (19)	193 (13)
Missing	4	14

	Ages 5–7 N = 432 (% of complete data)	Ages 8–17 N = 1,548 (% of complete data)
Child's age (yrs.)		
5–7	Mean = 6.0, SD = .83	---
8–12	---	665 (54)
13–17	---	566 (46)
Child's gender		
Male	189 (44)	736 (48)
Female	243 (56)	809 (52)
Missing	0	3

Table 2

Emotional health age DIF analysis

Short form	Item stem	Age DIF (5–7 vs. 8–17)		Age DIF (8–12 vs. 13–17)	
		X^2 (df)	p	X^2 (df)	p
	Depressive symptoms				
	My child wanted to be by himself/herself.	5.3 (5)	0.38	12.1 (5)	0.03
SF	My child felt alone.	4.3 (4)	0.36	4.1 (4)	0.40
SF	My child felt like he/she couldn't do anything right.	1.9 (4)	0.75	3.2 (4)	0.53
SF	My child felt everything in his/her life went wrong.	5.1 (4)	0.28	5.2 (4)	0.27
	My child felt too sad to eat.	2.7 (4)	0.62	2.6 (4)	0.62
	My child didn't care about anything.	4.3 (5)	0.51	10.1 (5)	0.07
	It was hard for my child to have fun.	5.0 (5)	0.42	6.4 (5)	0.27
SF	My child could not stop feeling sad.	4.1 (4)	0.39	2.6 (4)	0.64
SF	My child felt lonely.	10.0 (5)	0.08	9.1 (5)	0.11
SF	My child felt unhappy.	12.7 (5)	0.03	7.7 (5)	0.18
	Depressive symptoms DIF Items				
SF	My child felt sad. *	19.5 (5)	0.00	7.0 (5)	0.22
SF	My child thought that his/her life was bad. *	21.7 (5)	0.00	5.2 (5)	0.39
	It was hard for my child to do school work because he/she felt sad.	16.2 (5)	0.01	4.6 (5)	0.47
	My child felt stressed.	27.0 (5)	0.00	8.0 (5)	0.16
	Anxiety				
SF	My child felt scared.	5.3 (4)	0.26	3.0 (4)	0.55
SF	My child felt worried.	12.9 (5)	0.02	7.5 (5)	0.19
SF	My child felt like something awful might happen.	7.1 (5)	0.21	3.7 (5)	0.59
SF	My child worried about what could happen to him/her.	4.1 (5)	0.54	0.4 (5)	1.00
	My child worried when he/she was away from home.	9.5 (5)	0.09	7.6 (5)	0.18
SF	My child felt nervous.	9.9 (5)	0.08	4.6 (5)	0.47
	My child was worried he/she might die.	4.6 (3)	0.20	0.3 (3)	0.96
	It was hard for my child to relax.	10.7 (5)	0.06	8.2 (5)	0.15
	My child was afraid of going to school.	5.0 (5)	0.42	8.8 (5)	0.12
SF	My child was afraid that he/she would make mistakes.	8.2 (5)	0.15	7.2 (5)	0.21

Short form	Item stem	Age DIF (5–7 vs. 8–17)		Age DIF (8–12 vs. 13–17)	
		X^2 (df)	p	X^2 (df)	p
	Anxiety DIF items				
	My child worried when he/she was at home. *	13.1 (4)	0.01	4.7 (4)	0.32
SF	My child worried when he/she went to bed at night. *	19.2 (5)	0.00	1.1 (5)	0.96
SF	My child thought about scary things. *	14.2 (5)	0.01	11.0 (5)	0.05
	My child got scared really easy.	34.5 (5)	0.00	6.6 (5)	0.25
	My child woke up at night scared.	19.5 (4)	0.00	8.0 (4)	0.09
	Anger				
SF	My child felt mad.	1.2 (5)	0.94	6.0 (5)	0.31
SF	When my child got mad, he/she stayed mad.	1.6 (5)	0.90	6.2 (5)	0.29
SF	My child felt upset.	5.5 (4)	0.24	2.2 (4)	0.71
SF	My child was so angry he/she felt like throwing something.	2.6 (5)	0.76	5.9 (5)	0.32
SF	My child was so angry he/she felt like yelling at somebody.	4.8 (4)	0.31	7.9 (4)	0.09

*
Items that exhibited statistical DIF that were retained with caution.

Fatigue age DIF analysis

Table 3

Short form	Item stem	Age DIF (5–7 vs. 8–17)		Age DIF (8–12 vs. 13–17)	
		X^2 (df)	p	X^2 (df)	p
	Lack of energy				
SF	My child felt strong (not weak).	6.1 (5)	0.29	2.2 (5)	0.82
SF	My child had enough energy to do things outside.	4.4 (5)	0.50	8.3 (5)	0.14
	My child had enough energy to read.	3.5 (5)	0.62	10.9 (5)	0.05
	My child had enough energy to take a bath or shower.	11.1 (5)	0.05	11.4 (5)	0.04
SF	My child had enough energy to focus on his/her work.	**	**	**	**
SF	My child had enough energy to go out with his/her family.	**	**	**	**
SF	My child had enough energy to do sports or exercise.	**	**	**	**
SF	My child had energy.	**	**	**	**
SF	My child had enough energy to do the things he/she likes to do.	***	***	***	***
	Lack of energy DIF items				
SF	My child felt full of energy. *	16.1 (4)	0.00	2.7 (4)	0.60
SF	My child had enough energy to go out or play with his/her friends. *	20.9 (5)	0.00	1.2 (5)	0.94
	Tired				
SF	My child had trouble starting things because he/she was too tired.	11.4 (5)	0.04	1.8 (5)	0.87
	Being tired kept my child from having fun.	2.7 (3)	0.44	7.3 (3)	0.06
	Being tired made it hard for my child to keep up with schoolwork.	4.5 (5)	0.48	5.6 (5)	0.35
	My child was so tired it was hard for him/her to pay attention.	6.4 (5)	0.27	3.8 (5)	0.58
	My child had trouble finishing things because he/she was too tired.	8.2 (5)	0.15	10.2 (5)	0.07
	My child felt tired.	8.7 (4)	0.07	7.6 (4)	0.11
	My child was too tired to go out with his/her family.	9.4 (4)	0.05	2.0 (4)	0.74
SF	My child was too tired to do things outside.	8.4 (5)	0.13	4.0 (5)	0.55
	My child felt more tired than usual when he/she woke up in the morning.	6.5 (5)	0.26	2.6 (5)	0.76
	My child was too tired to read.	3.6 (5)	0.61	6.0 (5)	0.30
SF	My child was too tired to enjoy the things he/she likes to do.	**	**	**	**
SF	Being tired made it hard for my child to play or go out with friends as much as he/she would like.	**	**	**	**
SF	My child felt weak.	**	**	**	**

Short form	Item stem	Age DIF (5–7 vs. 8–17)		Age DIF (8–12 vs. 13–17)	
		X^2 (df)	p	X^2 (df)	p
SF	My child got tired easily.	**	**	**	**
	My child was too tired to watch television.	**	**	**	**
	My child was too tired to focus on his/her work.	**	**	**	**
	My child was too tired to go up and down a lot of stairs.	**	**	**	**
	It was hard for my child to get out of bed in the morning.	**	**	**	**
	My child was too tired to take a bath or shower.	**	**	**	**
	Tired DIF items				
SF	My child was too tired to do sports or exercise. *	17.6 (5)	0.00	5.5 (5)	0.36
	My child felt too tired to spend time with his/her friends. *	15.1 (4)	0.00	3.9 (4)	0.42
	My child was too tired to eat. *	12.7 (4)	0.01	5.1 (4)	0.27
	My child needed to sleep during the day.	16.8 (5)	0.00	22.3 (5)	0.00

*
Items that exhibited statistical DIF that were retained with caution.

**
Form 4 items are not included due to insufficient sample size.

***
Items are not included due to slope estimates larger than 10.

Table 4

Physical functioning DIF analysis

Short form	Item stem	Age DIF (5–7 vs. 8–17)		Age DIF (8–12 vs. 13–17)	
		X^2 (df)	p	X^2 (df)	p
	Upper extremity				
SF	My child could put on his/her shoes without help.	3.9 (5)	0.56	0.1 (5)	1.00
SF	My child could button his/her shirt or pants.	3.6 (5)	0.61	0.3 (5)	1.00
	My child could put on his/her clothes without help.	2.6 (5)	0.76	2.3 (5)	0.80
	My child could pull on and fasten his/her seatbelt.	3.6 (4)	0.47	2.0 (4)	0.73
	My child could put on his/her socks without help.	6.5 (5)	0.26	1.2 (5)	0.94
	My child could use a mouse or touch pad for the computer.	1.5 (3)	0.68	2.7 (3)	0.44
	My child could wash his/her face with a cloth.	3.0 (3)	0.40	1.8 (3)	0.61
SF	My child could use a key to unlock a door.	12.7 (5)	0.03	2.6 (5)	0.77
SF	My child could open the rings in school binders.	8.6 (5)	0.13	2.9 (5)	0.72
	My child could tie shoelaces without help.	6.4 (5)	0.27	2.6 (5)	0.77
	My child needed help with a bath.	8.6 (5)	0.13	3.8 (5)	0.58
SF	My child could pull open heavy doors.	9.0 (5)	0.11	1.7 (5)	0.88
	My child could move his/her hands or fingers.	3.4 (3)	0.33	4.0 (3)	0.26
	My child used a pencil with a special grip to write.	3.1 (5)	0.68	3.5 (5)	0.62
	My child could hold a full cup.	4.0 (3)	0.37	5.0 (3)	0.17
	My child could open his/her clothing drawers.	**	**	**	**
SF	My child could pour a drink from a full pitcher.	**	**	**	**
	My child could dry his/her back with a towel.	**	**	**	**
	My child could turn door handles without help.	**	**	**	**
SF	My child could open a jar by himself/herself.	**	**	**	**
	My child could brush his/her teeth without help.	**	**	**	**
	My child could write with a pen or pencil.	**	**	**	**
SF	My child could pull a shirt on over his/her head without help.	***	***	***	***
	My child could zip up his/her clothes.	***	***	***	***
	My child could put toothpaste on his/her toothbrush without help.	***	***	***	***
	My child could cut paper with scissors.	***	***	***	***

Short form	Item stem	Age DIF (5–7 vs. 8–17)		Age DIF (8–12 vs. 13–17)	
		X^2 (df)	p	X^2 (df)	p
	My child could lift a cup to drink.	***	***	***	***
	My child could hold an empty cup.	***	***	***	***
	Upper extremity DIF items				
	My child could dial a phone.	17.0 (4)	0.00	4.4 (4)	0.35
	Mobility				
SF	My child could stand up without help.	3.2 (3)	0.36	1.2 (3)	0.75
SF	My child could walk up stairs without holding on to anything.	3.4 (5)	0.64	5.4 (5)	0.37
SF	My child could stand up on his/her tiptoes.	1.4 (5)	0.92	3.7 (5)	0.60
	My child could get up from a regular toilet.†	9.1 (4)	0.06	36.7 (4)	0.00†
SF	My child could get down on his/her knees without holding on to something.	4.2 (5)	0.53	7.6 (5)	0.18
SF	My child could move his/her legs.	1.1 (4)	0.90	4.8 (4)	0.31
	My child could ride a bike.	0.9 (5)	0.97	3.7 (5)	0.59
	My child could bend over to pick something up.	**	**	**	**
	My child could get in and out of a car.	**	**	**	**
	My child could walk more than one block.	**	**	**	**
	My child used a wheelchair to get around.	**	**	**	**
	My child could get out of bed by himself/herself.	**	**	**	**
SF	My child could keep up when he/she played with other kids.	**	**	**	**
	My child could carry his/her books in a backpack.	**	**	**	**
	My child could get into bed by himself/herself.	***	***	***	***
	My child could walk across the room.	***	***	***	***
SF	My child could get up from the floor.	***	***	***	***
SF	My child has been physically able to do the activities he/she enjoys most.	***	***	***	***
	My child used a walker, cane or crutches to get around.	***	***	***	***
	My child could turn his/her head all the way to the side.	****	****	****	****
	Mobility DIF items				
SF	My child could do sports and exercise that other kids his/her age could do. *	17.3 (5)	0.00	2.0 (5)	0.85
	My child could go up one step. *	12.7 (3)	0.01	1.4 (3)	0.71
	My child could run a mile.	52.9 (5)	0.00	1.4 (5)	0.92

†Item exhibited DIF between higher age groups only.

*
Items that exhibited statistical DIF that were retained with caution.

**
Form 4 items are not included due to insufficient sample size.

***
Items are not included due to slope estimates larger than 10.

****
Item is not included due to only one observed response category.

Table 5

Pain interference age DIF analysis

Short form	Item stem	Age DIF (5–7 vs. 8–17)		Age DIF (8–12 vs. 13–17)	
		X^2 (df)	p	X^2 (df)	p
SF	It was hard for my child to have fun when he/she had pain.	5.4 (5)	0.37	4.9 (5)	0.42
SF	It was hard for my child to pay attention when he/she had pain.	4.3 (5)	0.51	5.4 (5)	0.36
SF	My child had trouble doing schoolwork when he/she had pain.	4.3 (5)	0.51	4.4 (5)	0.49
SF	My child had trouble sleeping when he/she had pain.	5.2 (5)	0.39	1.1 (5)	0.95
SF	It was hard for my child to run when he/she had pain.	10.6 (5)	0.06	5.3 (5)	0.39
SF	It was hard for my child to walk one block when he/she had pain.	7.6 (5)	0.18	10.0 (5)	0.08
SF	My child felt angry when he/she had pain.	8.9 (4)	0.06	1.7 (4)	0.79
	My child hurt all over his/her body.	6.0 (4)	0.20	3.6 (4)	0.47
	It was hard for my child to remember things when he/she had pain.	**	**	**	**
SF	It was hard for my child to stay standing when he/she had pain.	**	**	**	**
	It was hard for my child to get along with other people when he/she had pain.	**	**	**	**
	My child hurt a lot.	**	**	**	**
	My child missed school when he/she had pain.	**	**	**	**

**
Form 4 items are not included due to insufficient sample size.

Table 6

Peer Relationships age DIF analysis

Short form	Item stem	Age DIF (5–7 vs. 8–17)		Age DIF (8–12 vs. 13–17)	
		X^2 (df)	p	X^2 (df)	p
SF	Other kids wanted to be with my child.	3.6 (4)	0.46	7.0 (4)	0.14
	My child felt good about his/her friendships.	2.0 (4)	0.74	4.2 (4)	0.38
	My child was able to have fun with his/her friends.	3.1 (4)	0.54	2.6 (4)	0.63
SF	Other kids wanted to talk to my child.	8.2 (5)	0.14	7.1 (5)	0.22
SF	My child was good at making friends.	2.2 (4)	0.70	5.0 (4)	0.29
SF	My child felt accepted by other kids his/her age.	2.8 (5)	0.17	7.8 (5)	0.17
SF	My child and his/her friends helped each other out.	1.2 (5)	0.95	5.1 (5)	0.40
	My child was a good friend.	3.2 (4)	0.53	3.4 (4)	0.49
	My child shared with other kids (food, games, pens, etc.).	5.2 (5)	0.39	1.1 (5)	0.96
SF	My child was able to count on his/her friends.	**	**	**	**
	My child liked being around other kids his/her age.	**	**	**	**
SF	My child was able to talk about everything with his/her friends.	**	**	**	**
	My child spent time with his/her friends.	**	**	**	**
	My child played alone and kept to himself/herself.	**	**	**	**
	DIF items				
SF	Other kids wanted to be my child's friend. *	30.1 (3)	0.00	3.0 (3)	0.39

*
Items that exhibited statistical DIF that were retained with caution.

**
Form 4 items are not included due to insufficient sample size.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Asthma age DIF analysis

Table 7

Short form	Item stem	Age DIF (5–7 vs. 8–17)		Age DIF (8–12 vs. 13–17)	
		X^2 (df)	p	X^2 (df)	P
SF	My child's asthma bothered him/her.	3.9 (5)	0.57	1.4 (5)	0.92
SF	My child had trouble breathing because of his/her asthma.	1.0 (5)	0.96	3.5 (5)	0.62
SF	My child felt wheezy because of his/her asthma.	2.6 (5)	0.77	1.8 (5)	0.88
SF	It was hard for my child to take a deep breath because of asthma.	4.2 (5)	0.53	3.8 (5)	0.58
	My child was bothered by the amount of time he/she spent wheezing.	6.4 (4)	0.17	3.0 (4)	0.57
	My child had asthma attacks.	0.8 (5)	0.98	4.1 (5)	0.53
SF	My child's chest felt tight because of asthma.	7.0 (5)	0.22	3.1 (5)	0.69
	My child's body felt bad when he/she was out of breath.	2.6 (5)	0.76	1.6 (5)	0.90
SF	My child had trouble sleeping at night because of asthma.	5.3 (5)	0.38	6.1 (5)	0.30
	My child was bothered by asthma when he/she was with friends.	0.6 (4)	0.96	3.1 (4)	0.54
	My child got tired easily because of his/her asthma.	3.6 (4)	0.46	2.0 (4)	0.73
SF	It was hard for my child to play sports or exercise because of asthma.	5.0 (5)	0.41	2.4 (5)	0.79
	My child had trouble walking because of asthma.	7.6 (4)	0.11	4.7 (4)	0.32
SF	My child felt scared that he/she might have trouble breathing because of asthma.	1.5 (5)	0.91	1.6 (5)	0.90
	My child missed school because of asthma.	9.8 (4)	0.04	3.8 (4)	0.44
	It was hard for my child to play with pets because of asthma.	3.2 (5)	0.67	5.1 (5)	0.40
	DIF items				
	My child coughed because of his/her asthma. *	19.6 (5)	0.00	1.3 (5)	0.94

*
Items that exhibited statistical DIF that were retained with caution.