

11-1-2016

## Differential item functioning in the Patient Reported Outcomes Measurement Information System Pediatric Short Forms in a sample of children and adolescents with cerebral palsy.

Wendy J. Coster  
*Boston University*

Pengsheng Ni  
*Boston University*

Mary D. Slavin  
*Boston University*

Pamela A. Kisala  
*University of Delaware*

Follow this and additional works at: <https://jdc.jefferson.edu/otfp>

Ratna Nandakumar

 *University of Delaware* Part of the [Occupational Therapy Commons](#), and the [Pediatrics Commons](#)

### [Let us know how access to this document benefits you](#)

*See next page for additional authors*

#### Recommended Citation

Coster, Wendy J.; Ni, Pengsheng; Slavin, Mary D.; Kisala, Pamela A.; Nandakumar, Ratna; Mulcahey, Mary Jane; Tulskey, David S.; and Jette, Alan M., "Differential item functioning in the Patient Reported Outcomes Measurement Information System Pediatric Short Forms in a sample of children and adolescents with cerebral palsy." (2016). *Department of Occupational Therapy Faculty Papers*. Paper 61.  
<https://jdc.jefferson.edu/otfp/61>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Department of Occupational Therapy Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: [JeffersonDigitalCommons@jefferson.edu](mailto:JeffersonDigitalCommons@jefferson.edu).

---

## Authors

Wendy J. Coster, Pengsheng Ni, Mary D. Slavin, Pamela A. Kisala, Ratna Nandakumar, Mary Jane Mulcahey, David S. Tulsy, and Alan M. Jette



Published in final edited form as:

*Dev Med Child Neurol*. 2016 November ; 58(11): 1132–1138. doi:10.1111/dmcn.13138.

## Differential item functioning in the Patient Reported Outcomes Measurement Information System Pediatric Short Forms in a sample of children and adolescents with cerebral palsy

Wendy J Coster<sup>1</sup>, Pengsheng Ni<sup>2</sup>, Mary D Slavin<sup>2</sup>, Pamela A Kisala<sup>3</sup>, Ratna Nandakumar<sup>4</sup>, Mary Jane Mulcahey<sup>5</sup>, David S Tulskey<sup>3,6</sup>, and Alan M Jette<sup>2</sup>

<sup>1</sup>Department of Occupational Therapy, Boston University College of Health and Rehabilitation Sciences: Sargent College, Boston, MA

<sup>2</sup>Health and Disability Research Institute, Boston University School of Public Health, Boston, MA

<sup>3</sup>Department of Physical Therapy, College of Health Sciences, University of Delaware, Newark, DE

<sup>4</sup>College of Education and Human Development, University of Delaware, Newark, DE

<sup>5</sup>Department of Occupational Therapy, Thomas Jefferson University, Philadelphia, PA

<sup>6</sup>Kessler Foundation Research Center, West Orange, NJ, USA

### Abstract

**AIM**—The present study examined the Patient Reported Outcomes Measurement Information System (PROMIS) Mobility, Fatigue, and Pain Interference Short Forms (SFs) in children and adolescents with cerebral palsy (CP) for the presence of differential item functioning (DIF) relative to the original calibration sample.

**METHOD**—Using the Graded Response Model we compared item parameter estimates generated from a sample of 303 children and adolescents with CP (175 males, 128 females; mean age 15y 5mo) to parameter estimates from the PROMIS calibration sample, which served as the reference group. DIF was assessed in a two-step process using the item response theory–likelihood ratio–differential item functioning detection procedure.

**RESULTS**—Significant DIF was identified for four of eight items in the PROMIS Mobility SF, for two of eight items in the Pain Interference Scale, and for one item out of 10 on the Fatigue

---

Correspondence to Wendy Coster, Department of Occupational Therapy, Boston University, College of Health and Rehabilitation Sciences: Sargent College, 635 Commonwealth Avenue, Boston, MA 02215, USA. wjcoster@bu.edu.

The authors have stated that they had no interests which might be perceived as posing a conflict or bias.

### SUPPORTING INFORMATION

The following additional material may be found online:

**Table SI:** Mobility Short Form: item-level descriptive statistics.

**Table SII:** Mobility Short Form: S- $\chi^2$  item-level diagnostic statistics.

**Table SIII:** Fatigue Short Form: item-level descriptive statistics.

**Table SIV:** Fatigue Short Form: S- $\chi^2$  item-level diagnostic statistics.

**Table SV:** Pain Interference Short Form: item-level descriptive statistics.

**Table SVI:** Pain Interference SF: S- $\chi^2$  item-level diagnostic statistics.

Scale. Impact of DIF on total score estimation was notable for Mobility and Pain Interference, but not for Fatigue.

**INTERPRETATION**—Results suggest differences in the responses of adolescents with CP to some items on the PROMIS Mobility and Pain Interference SFs. Cognitive interviews about the PROMIS items with adolescents with varying degrees of mobility limitations would provide better understanding of how they are interpreting and selecting responses to the PROMIS items and thus help guide selection of the most appropriate way to address this issue.

The Patient Reported Outcomes Measurement Information System (PROMIS) was developed to support measurement of health-related quality of life for children, adolescents, and adults. Each domain consists of item response theory (IRT) calibrated item banks, in which items are hierarchically organized from low to high levels of the respective ability or symptom experience. PROMIS measures may be administered either as computerized adaptive tests or as brief, fixed-length Short Forms (SFs) consisting of eight to 10 items that span the continuum of the latent trait being assessed.

PROMIS instruments use an approach to measurement that is domain-specific, rather than disease- or condition-specific.<sup>1</sup> Items were designed to capture the common components of a concept on the assumption that these are relevant regardless of one's condition. This approach rests on the expectation that the items in the instrument measure the same underlying trait (latent trait) regardless of the patient population responding to them. However the PROMIS developers also acknowledge the need for continuing validation research to examine whether the measures perform equally well in different clinical groups, thus yielding a valid and generalizable assessment from a single, general set of item parameters.<sup>1, 2</sup> Two related validity issues need to be considered. First, are the items equally appropriate for all clinical groups? Second, do those items show measurement invariance? Measurement invariance refers to the expectation that the likelihood of an item response from two people at the same level of the latent trait does not differ depending on their group membership. This issue is especially relevant in clinical populations where the measurement construct reflects an area of primary impairment or symptom experience. These impairments or experiences may alter the difficulty of the activities relative to one another, and therefore lead to different item performance for participants with the same ability level on the underlying trait, resulting in differential item functioning (DIF).

The PROMIS pediatric Mobility, Fatigue, and Pain Interference item banks were calibrated in a sample of more than 4000 children, of whom 28% reported at least one chronic medical condition.<sup>3–5</sup> However, the most common conditions were asthma (18% of sample), attention-deficit disorder/attention-deficit–hyperactivity disorder (4.6%), and arthritis, gastrointestinal disorders, mental disorders, immune disorders, and allergies (1–3% each).<sup>4</sup> There were few participants with physically-disabling chronic conditions such as cerebral palsy (CP). Because the impact of CP on daily life may be quite different, it is important to evaluate measurement invariance in this population. The present study compared item parameters for the Mobility, Fatigue, and Pain Interference SF items derived from PROMIS calibration samples to item parameter estimates from children and adolescents with CP. Specific research questions were as follows. Are the PROMIS pediatric SF item parameter

estimates for these SFs equivalent across the PROMIS calibration sample and children and adolescents with CP? If item parameter estimates differ, how much do these differences affect IRT-based summary scores for children and adolescents with CP?

## METHOD

### Participants

Participants with CP were a combined sample of children and adolescents who had completed the three SFs in two studies described below.

The first CP sample ( $n=113$ ) was from a study examining responsiveness of PROMIS measures following elective musculoskeletal surgery.<sup>6</sup> PROMIS SF data from the baseline assessment (pre-surgery) of a sample (aged 8–21y) with a confirmed diagnosis of CP was used. Exclusion criteria were: orthopedic surgery for reason other than to improve physical functioning; cognitive impairment that limited the child's ability to read, understand, and respond to items; primary language other than English; and functional limitations caused by meningitis, brain tumors, or acquired injuries, and/or disease of the brain.

The second CP sample ( $n=190$ ) was obtained from a study of children, adolescents, and young adults (aged 14–25y) to develop a scoring link between pediatric and adult PROMIS item banks.<sup>7</sup> Inclusion criteria included the ability to understand and respond (speaking, using a communication board, or gesturing) to self-report questions, and CP diagnosis confirmed through medical record review. Each site participating in these two studies obtained local Institutional Review Board approval.

PROMIS sample calibration estimates were obtained from the PROMIS project. Estimates were derived from a large-scale study conducted in two states, North Carolina and Texas, described in detail in previous publications.<sup>4, 5</sup> Participants aged 8 to 17 years were recruited from outpatient pediatric and specialty clinics, and public school settings. Participants (including the consenting parent or guardian) needed to be able to speak and read English and the adolescent needed to be able to see and interact with a computer screen, keyboard, and mouse.

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

### Procedure

Participants aged 18 years or older provided informed consent. Participants in sample 1 used a tablet computer to complete the PROMIS measures via the Assessment Center platform (Department of Medical Social Sciences, Northwestern University, Chicago, IL).<sup>8</sup> A study team member remained in the room with the participant in case any questions or problems arose. Sample 2 participants completed the measures in interview format, either in person or over the phone. Color-coded response cards were provided to ensure that participants were referencing the appropriate response set for each item, and all responses were entered

directly into the Assessment Center platform. All data collectors completed a standardized webinar-based training process and followed a detailed manual of procedures.

Participants in the PROMIS calibration study completed the instruments on laptop computers without parent or peer assistance. Each child completed one of seven different administration forms, thus different numbers of children completed particular items within each domain.

## Instruments

All participants completed the Mobility (eight items), Pain Interference (eight items), and Fatigue (10 items) SFs. The measures use a common context: 'In the past 7 days...' followed by a specific item stem and five response options: 'with no trouble', 'a little trouble', 'some trouble', 'a lot of trouble', 'unable to do' (Mobility), where higher scores indicate better mobility function; and 'never', 'almost never', 'sometimes', 'often', 'almost always' (Fatigue; Pain Interference), where higher scores indicate more pain or fatigue. All measures were normed on the PROMIS pediatric calibration samples and are scored on a T metric with a mean of 50 and standard deviation (SD) of 10.

The Gross Motor Function Classification System – Expanded and Revised version (GMFCS – E&R)<sup>9, 10</sup> was used to describe the severity of participants' mobility limitations.

## Data analysis

The marginal maximum likelihood estimation method was applied using IRTPRO (Scientific Software International Inc., Skokie, IL) to calibrate the items. We selected the Graded Response Model to fit the items, which has been commonly used in PROMIS calibration. The Graded Response Model uses a slope parameter and a number of threshold parameters to characterize the item response. The slope parameter (discrimination parameter) describes the strength of association between the item response and the latent trait. Because all items were recoded with the same direction (higher score indicates greater function), a higher slope in this study indicates that the item can discriminate more clearly between respondents with more or less of the trait. The threshold parameter represents the value on the latent trait where a participant will have 50% probability of responding with this rating or higher.

Before conducting DIF analyses, we applied confirmatory factor analysis to evaluate whether items in each scale fit a unidimensional model in the CP sample.

DIF analyses were conducted using the methodology outlined in other PROMIS studies.<sup>11</sup> To assess DIF across the two groups, we applied a two-step IRT-based method using the PROMIS calibration sample data as the reference group and CP data as the focal group. In the first step, we constrained the mean and SD of the PROMIS sample by setting them to 0 and 1 respectively. We then set the item parameters to be equal across the two samples to estimate the mean and SD of the CP sample. We next fixed the CP sample mean and SD and free estimated the item parameters across the two samples. The Wald  $\chi^2$  test was used to examine the equality of item parameters (including discrimination and threshold) across groups and identify DIF items. Because this method for DIF detection requires multiple tests of significance, we applied the Benjamini–Hochberg adjustment<sup>6, 12</sup> to control for Type 1

error in multiple comparisons. Items without DIF identified in this step were used as anchor items in the next step. In the second step, we constrained the non-DIF item parameters to be equal across the two samples, free estimated the mean and SD for the CP sample and the parameters of the rest of the items, and examined DIF using a two-group IRT model.<sup>13, 14</sup>

Graphical methods were used to evaluate the magnitude of effect sizes when significant DIF was detected.<sup>13</sup> For items that exhibited DIF (i.e. parameters were significantly different), we visually examined Item Characteristic Curves (ICC) and Option Characteristic Curves to identify trends. First, we examined whether the DIF was uniform or non-uniform. For uniform DIF, we evaluated whether the item was easier or harder in the CP sample compared to the PROMIS sample. For non-uniform DIF we identified the score range in which the item became easier or harder, and trends of items that were easier or harder relative to the PROMIS estimates. We determined the DIF impact by calculating the weighted area between the expected score curves (wABC).<sup>15</sup> For items that include five response options, wABC greater than 0.3 suggests possible problematic DIF and need for further investigation. Finally, to examine DIF impact at the score level, we calculated two summed scores based on the IRT score: one based on the item parameters generated from the sample; and the other based on the item parameters generated from the CP sample. Item parameters were set to be equal in DIF-free items. In calculating the IRT score, the CP sample distribution was set as the target population.

We also examined DIF by sex and age groups (<18y; ≥18y).

## RESULTS

Compared to the PROMIS calibration sample, the sample with CP included a higher percentage of males and had a higher mean age. Children and adolescents in the CP sample had a range of mobility limitations, with a greater representation at higher levels of functioning according to GMFCS. More than 80% of the participants with CP were ambulatory, although with some degree of limitation, and 16% of this group used some form of device when walking (Table I for details of both samples).

Results of confirmatory factor analysis supported the unidimensionality of the three scales in the CP sample: Fatigue,  $M^2=1050.25$ , root mean square error of approximation (RMSEA)=0.04; Pain Interference,  $M^2=806.99$ , RMSEA=0.05; and Mobility,  $M^2=638.6$ , RMSEA=0.04.

There was no DIF by either age group or sex in the CP sample.

### Mobility SF

Mean scores of the CP sample on individual Mobility SF items were consistently lower than item means in the PROMIS calibration sample, although the size of the difference varied (range=0.08–1.43).

The first step of the DIF analysis revealed six out of eight items with significant DIF. In step 2, parameters for the items without DIF were constrained to be equivalent, and the parameter estimates were again compared. Results showed that all six items continued to display

significant DIF. Among the six statistically significant DIF items, there were four items with a wABC value (weighted by the CP sample distribution) that exceeded the 0.3 criterion (Table II). If the PROMIS Mobility SF item parameters were used to score the CP sample, mobility would be overestimated in the lower summed score range and underestimated (but less so) in higher summed score ranges (Fig. 1). The impact would be more severe in the lower range (the range where the participants with CP tend to score) than in the higher range. The largest score difference (0.3 logit) corresponded to a medium effect size (0.25). Inspection of ICCs identified one item with a particularly notable difference, 'I have been physically able to do the activities I enjoy most'. The participants with CP gave much higher ratings on this item than participants in the calibration sample at comparable levels of overall estimated ability. (Fig. 2).

### Pain interference

Individual item mean scores on the Pain Interference SF varied across the samples (range of differences = 0.02–0.28), with some means higher and some lower in the CP sample compared to the calibration sample.

DIF analyses (Table II) revealed five items with significant DIF, however only two of the items also had wABC that exceeded the criteria 'It was hard for me to walk one block when I had pain' and 'It was hard to stay standing when I had pain' (wABC=0.45 for both). For both items, inspection of ICCs revealed that participants with CP gave higher ratings on the item, indicating greater perceived limitation, than participants in the calibration sample with the same overall level of pain interference. Comparison of the summed scores calculated using PROMIS parameters versus CP parameters showed score differences across most of the range with the largest difference (approximately 0.35 logits) at summed score 2.

### Fatigue SF

The individual item means of the two groups were more similar in this domain (range of difference=0–0.17). Only one item had significant DIF and a relatively large wABC (0.7; Table II). The impact of this one item on overall differences in summed scores was quite small, i.e. summary score curves were almost identical across the full range of scores. Additional details of the analyses are provided in Tables SI–SVI (online supplementary information).

## DISCUSSION

Documentation of intervention outcomes of children and adolescents with CP requires measures that accurately represent the experience of the young people. The present study examined one aspect of this validity question for the PROMIS Mobility, Fatigue, and Pain Interference SF measures. IRT-based comparison of item calibration estimates from the original PROMIS sample to calibration estimates from a sample of children and adolescents with CP revealed that some items performed differently (i.e. had significant DIF). DIF was most common in mobility, where four of the eight items had significantly different parameter estimates. Findings suggested that if PROMIS item parameters were used, the



summary score might overestimate the ability (perceived mobility latent trait) of children and adolescents with CP compared to young people without CP who obtain the same score.

There are several possible explanations for these differences. The first is that the perceived difficulty of certain items is greater for children and adolescents with CP because of their underlying movement disorder. That is, the unique impact of the movement disorder might alter the relative degree of difficulty in performing the particular activities described in the items that had DIF. This explanation fits data from two of the mobility items (stand on tip-toes; go upstairs without holding on), where the ICCs show that the participants with CP tended to give lower ratings than participants in the PROMIS calibration sample with the same overall ability level. A second explanation is that the children and adolescents with CP interpret and respond differently to certain items. This could explain why the participants with CP gave higher ratings to two of the items with DIF ('keep up when playing'; 'physically able to do activities I enjoy most') compared to PROMIS calibration sample participants at the same ability level. In contrast to health conditions that have a more episodic impact on a child's mobility (e.g. asthma), the impact of CP on the child's mobility is consistent. Thus, children and adolescents with CP responding to an item such as 'do activities I enjoy most' are likely to reflect on the activities they typically do, i.e. that are within their range of ability, not on activities that are outside this range. In this case, the response reflects a different interpretation of the item compared with children and adolescents in the calibration sample. Indications that children and adolescents with CP were interpreting and responding differently were also described by Kratz et al.<sup>16</sup> who analyzed performance of the PROMIS Mobility computerized adaptive tests in a subsample of the current sample with CP. These two explanations are not mutually exclusive; it is possible that both may affect responses to the items.

Different interpretation may also explain the two items with DIF on the Pain Interference scales ('hard to stay standing'; 'hard to walk 1 block'). On both items, the adolescent with CP gave higher ratings (i.e. reported more pain interference) than PROMIS participants with the same overall level of pain interference. Approximately 60% of the participants with CP either do not walk or have difficulty walking because of their motor impairment, which raises the question of whether their response to these items specifically reflects their degree of limitation caused by pain, as the scale intended, or their overall degree of movement limitation. These possible differences in perspective are relevant if one wants to use the PROMIS SFs to compare mobility and pain interference across populations because equivalent scores will not have the same meaning.

In contrast to results from the Mobility and Pain Interference SFs, only one item showed DIF on the Fatigue SF, and its impact on total score was negligible. It is possible that children and adolescents with CP are better able to distinguish fatigue as a transient state and therefore evaluate its impact on their recent daily activities in a way that is more similar to that of the PROMIS calibration sample. In addition, unlike the Mobility and Pain Interference SFs, the Fatigue items do not ask about impact of fatigue on discrete activities. The one movement-related item ('do sports and exercise') is quite broad.

There are several options to address these measurement issues. First, cognitive interviews of the PROMIS SF items could help determine whether children and adolescents with CP with different degrees of mobility limitation or pain are interpreting items and response options in a similar way to participants in previous PROMIS studies. If so, results would support applying DIF adjustments to the Mobility and Pain Interference SFs using equating procedures to enable comparison across populations using the same forms and maintaining the same metric. This approach has been applied successfully elsewhere.<sup>17</sup> However, if the cognitive interviews suggest that some other factor – such as lack of understanding or a different interpretation of meaning, difficulty in making the comparison implied in the item, or impact of other clinical factors such as vision or upper extremity limitations – is influencing responses, this would indicate that the responses of children and adolescents with CP are not reflecting the same construct, which is a validity issue.

One option to address validity concerns would be to develop a modified SF that is specific to CP, following a process similar to that used to develop a Fatigue SF for persons with multiple sclerosis.<sup>18</sup> This process involves selecting items in the larger PROMIS item bank that best represent relevant activities or concerns for young people with CP, while also offering broad content and measurement coverage. This approach should incorporate cognitive interviews with children and adolescents with CP with varying degrees of movement limitations to examine how they interpret the potential items and select their responses. Research to evaluate the psychometric properties of a CP-specific PROMIS form with a new sample of participants with CP would be necessary.

If cognitive testing results did not support equating, or if review of the full item bank did not identify sufficient relevant items, another option would be to replenish the PROMIS item bank with additional items designed specifically to capture the experience of children and adolescents with CP. A modified scale with new and existing items would need to be calibrated and then advanced IRT linking techniques could be used to transform the CP calibrations to the PROMIS metric to allow direct comparison of scores.

## Limitations

Findings should be interpreted in relation to study limitations. The sample with CP had a larger percentage of males and a smaller proportion of younger participants than the PROMIS calibration samples, and included 107 individuals who were above the age range represented in the PROMIS sample. Previous studies during the development of these SFs reported some DIF by age and sex,<sup>9</sup> however most of those items were not included in the SFs and we did not find DIF by age or sex in our sample of children and adolescents with CP. Different modes of administration were used for the two CP samples, which may have contributed to measurement variance or score differences. Although different modes of administration did not affect scores for adult PROMIS mobility items,<sup>19</sup> this may or may not be the case for children and adolescents with CP.

## CONCLUSION

It is not clear whether summary scores on the PROMIS Mobility and Pain Interference SFs have the same meaning for children and adolescents with CP as for participants in the

PROMIS calibration sample. These results suggest that cognitive testing with current items would be useful to clarify whether the items in these SFs are measuring the same construct as intended by the developers or whether some revisions may be needed to the measures for the population with CP.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The study was funded through the Patient Reported Outcomes Measurement Information System (PROMIS) (National Institutes of Health grant # U01AR057929) and The Shriners Hospitals for Children Grant 79120. Boston and Cincinnati Children's Hospitals and The Shriners Hospitals for Children are acknowledged for their contributions to the work. David Thissen and Darren DeWalt from the PROMIS group at the University of North Carolina provided access to the PROMIS calibration data used in this study.

## ABBREVIATIONS

<b>DIF</b>	Differential item functioning
<b>IRT</b>	Item response theory
<b>PROMIS</b>	Patient Reported Outcomes Measurement Information System
<b>SF</b>	Short Form
<b>wABC</b>	Expected score curves

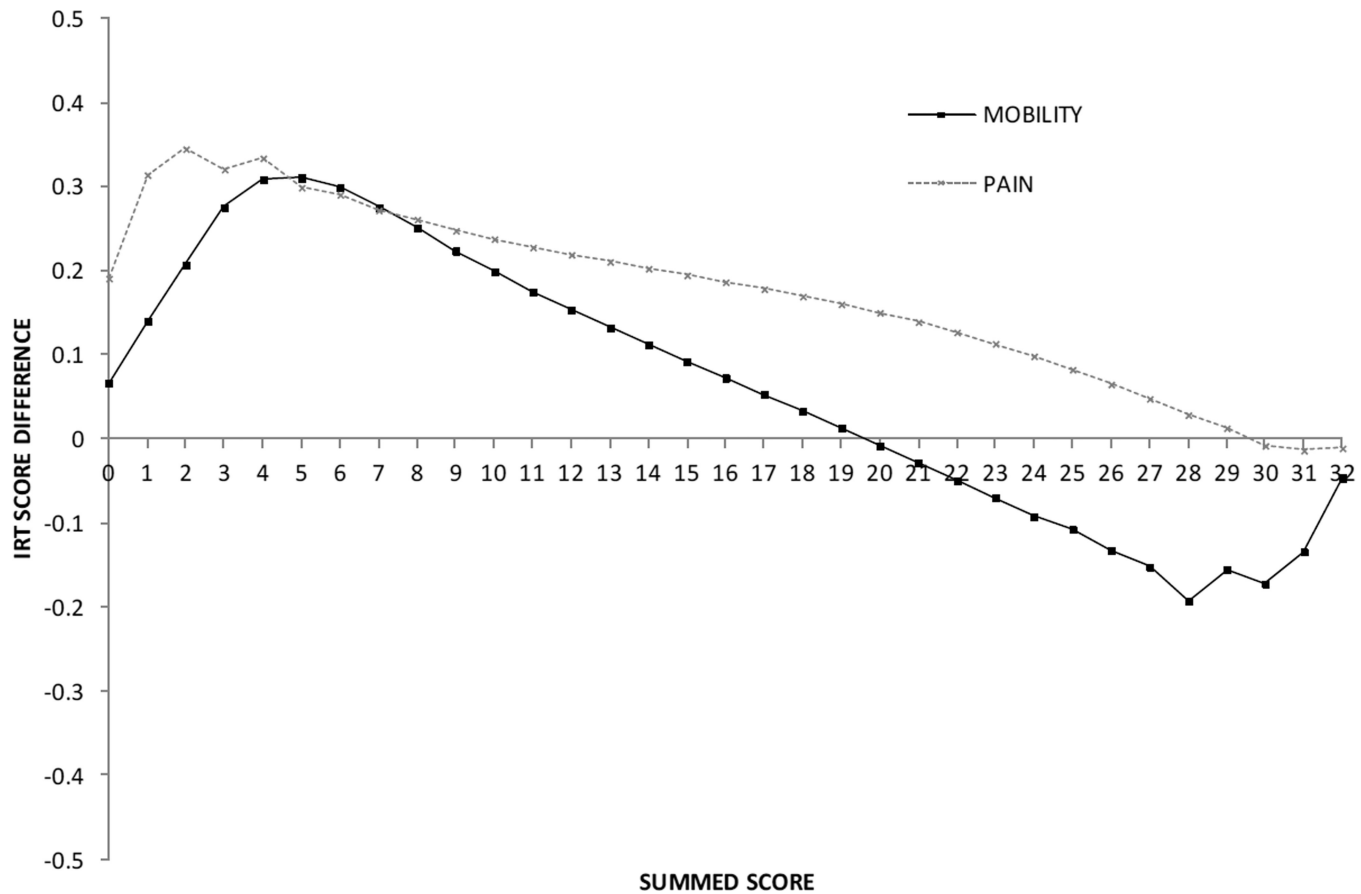
## REFERENCES

1. Magasi S, Ryan G, Revicki D, et al. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. *Qual Life Res.* 2012; 21:739–746. [PubMed: 21866374]
2. Cook, K.; Kallen, M.; Cella, D., et al. The Patient Reported Outcomes Measurement Information System (PROMIS®) perspective on: universally-relevant vs. disease-attributed scales. Chapel Hill, NC: PROMIS Network; 2014.
3. Lai J, Stucky B, Thissen D, et al. Development and psychometric properties of the PROMIS pediatric fatigue item banks. *Qual Life Res.* 2013; 22:2417–2427. [PubMed: 23378106]
4. Irwin D, Stucky B, Thissen D, et al. Sampling plan and patient characteristics of the PROMIS pediatrics large-scale survey. *Qual Life Res.* 2010; 19:585–594. [PubMed: 20204706]
5. Dewitt E, Stucky B, Thissen D, et al. Construction of the eight-item patient-reported outcomes measurement information system pediatric physical function scales: built using item response theory. *J Clin Epidemiol.* 2011; 64:794–804. [PubMed: 21292444]
6. Mulcahey M, Slavin M, Ni P, Kisala PA, Ni P, Tulskey DS, Jette AM. Ability of PROMIS pediatric measures to detect change in children with cerebral palsy undergoing musculoskeletal surgery. *J Pediatr Orthop.* 2015 Jun 5. [E-pub ahead of print].
7. Reeve B, Thissen D, DeWalt D, et al. Linkage between the PROMIS pediatric and adult Emotional Distress measures. *Qual Life Res.* 2015 Sep 30. [Epub ahead of print].
8. Gershon R, Rothrock NE, Hanrahan RT, Jansky LJ, Harniss M, Riley W. The development of a clinical outcomes survey research application: Assessment Center. *Qual Life Res.* 2010; 19:677–685. [PubMed: 20306332]
9. Palisano, R.; Rosenbaum, P.; Bartlett, D., et al. GMFCS – E&R: Gross Motor Function Classification System Expanded and Revised. Hamilton, Ontario: CanChild, McMaster University; 2007.

10. Russell D, Avery L, Rosenbaum P, et al. Improved scaling of the gross motor function measure for children with cerebral palsy: evidence of reliability and validity. *Phys Ther.* 2000; 80:873–885. [PubMed: 10960935]
11. Varni J, Stucky B, Thissen D, et al. PROMIS Pediatric Pain Interference Scale: an item response theory analysis of the Pediatric Pain Item Bank. *J Pain.* 2010; 11:1109–1119. [PubMed: 20627819]
12. Cook K, Bamer A, Roddey T, Kraft GH, Kim J, Amtmann D. A PROMIS fatigue short form for use by individuals who have multiple sclerosis. *Qual Life Res.* 2012; 21:1021–1030. [PubMed: 21927914]
13. Woods C, Cai L, Wang M. The Langer-improved Wald test for DIF testing with multiple groups: evaluation and comparison to two-Group IRT. *Educ Psychol Meas.* 2013; 73:532–547.
14. Edelen M, Stucky B, Chandra A. Quantifying 'problematic' DIF within an IRT framework: application to a cancer stigma index. *Qual Life Res.* 2015; 24:95–103. [PubMed: 24214177]
15. Cai, L.; Thissen, D.; du Toit, S. IRTPRO: Flexible, Multidimensional, Multiple Categorical IRT Modeling [Computer software]. Lincolnwood, IL: Scientific Software International; 2011.
16. Kratz A, Slavin M, Mulcahey M, Jette AM, Tulskey DS, Haley SM. An examination of the PROMIS pediatric instruments to assess mobility in children with cerebral palsy. *Qual Life Res.* 2013; 22:2865–2876. [PubMed: 23543391]
17. Coster W, Kramer J, Tian F, Dooley M, Liljenquist K, Kao YC, Ni P. Evaluating the appropriateness of a new computer-administered measure of adaptive function for children and youth with autism spectrum disorders. *Autism.* 2016; 20:14–25. [PubMed: 25630376]
18. Cook K, Barner A, Roddey T, Kraft GH, Kim J, Amtmann D. A PROMIS fatigue short form for use by individuals who have multiple sclerosis. *Qual Life Res.* 2012; 21:1021–1030. [PubMed: 21927914]
19. Bjorner J, Rosed M, Gandek B, Stone AA, Junghaenel DU, Ware JE Jr. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. *J Clin Epidemiol.* 2014; 67:108–113. [PubMed: 24262772]

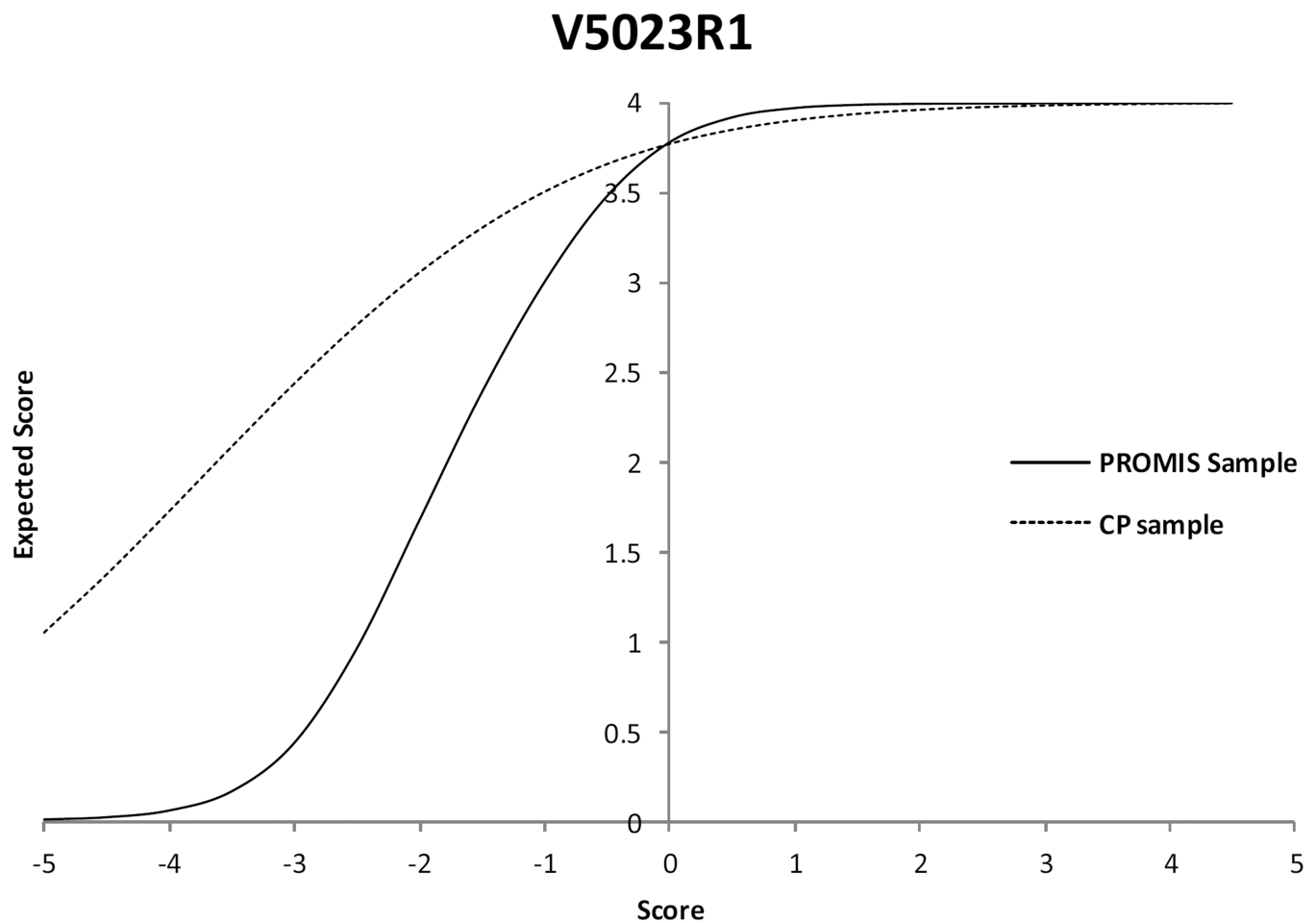
**What this paper adds**

- Children and adolescents with CP may not interpret some items on the PROMIS Mobility and Pain Interference Short Forms in the same way as children and adolescents in the original calibration sample.
- These differences could lead to overestimation of mobility function or pain interference for children and adolescents with CP.



**Figure 1.**

Score differences for Mobility and Pain Interference Short Form summed scores calculated with Patient Reported Outcomes Measurement Information System versus cerebral palsy-specific parameters. IRT, item response theory.



**Figure 2.**

Item characteristic curves comparison for the Mobility item 'I have been physically able to do the activities I enjoy most.' Participants in the lower range of the cerebral palsy sample gave higher ratings than participants in the Patient Reported Outcomes Measurement Information System sample at the same level of overall ability. PROMIS, Patient Reported Outcomes Measurement Information System; CP, cerebral palsy.

**Table I**

Demographic characteristics of samples: number and percentage

	PROMIS calibration samples			CP sample
	Mobility items <i>n</i> (%) <sup>a</sup>	Pain Interference items <i>n</i> (%) <sup>a</sup>	Fatigue items <i>n</i> (%) <sup>a</sup>	<i>n</i> =303
Male	2783 (47)	2708 (47)	2164 (47)	175 (58)
Female	3109 (53)	3048 (53)	2455 (53)	125 (42)
Missing	8	10	7	1
Age (mean, SD)	12.53 (2.7)	12.32 (2.7)	12.35 (2.7)	15.45 (3.2)
8–12	3102 (53)	3059 (53)	2426 (53)	51 <sup>b</sup> (17)
13–17	2785 (47)	2694 (47)	2189 (47)	148 (49)
18–25	0	0	0	107 (35)
Missing	13	13	11	1
GMFCS level <sup>c</sup>	NA			
Level I				105 (35)
Level II				96 (32)
Level III				49 (17)
Level IV				31 (10)
Level V				5 (2)
Unknown/missing				17 (6)

<sup>a</sup>Number who responded to one or more of the items included in the Short Form.<sup>b</sup>Children younger than 8 years: Mobility and Pain Interference samples (*n*=2); Fatigue sample (*n*=3).<sup>c</sup>Level I, walks without limitations; II, walks with limitations; III, walks using a handheld mobility device; IV, involves wheelchair self-mobility with limitations, may use power wheelchair; V, is transported in a manual wheelchair. PROMIS, Patient Reported Outcomes Measurement Information System; CP, cerebral palsy; SD, standard deviation; GMFCS, Gross Motor Function Classification System; NA, not available.



**Table II**

Results of step 2 differential item functioning analyses

Item: In the past 7 days...	$\chi^2$	df	Criterion $p^a$	p	wABC
<b>Mobility SF</b>					
I could get up from the floor.	21.9	5	0.025	0.0005	0.113
I could do sports and exercise that other kids my age could do.	26.4	5	0.021	0.0001	0.213
I could keep up when I played with other kids.	45.5	5	0.017	0.0001	0.308
I could stand up on my tiptoes.	51.9	5	0.013	0.0001	0.452
I could walk up stairs without holding on to anything.	59.7	5	0.008	0.0001	0.588
I have been physically able to do the activities I enjoy most.	127.8	5	0.004	0.0001	0.660
<i>Anchor Items</i>					
<i>I could move my legs</i>					
<i>I could stand up by myself</i>					
<b>Pain Interference SF</b>					
I felt angry when I had pain.	n.s.				
It was hard to have fun when I had pain.	19.6	5	0.014	0.0015 <sup>a</sup>	0.164
I had trouble doing schoolwork when I had pain.	10.9	5	0.025	0.0535	n.s.
It was hard for me to run when I had pain.	45.4	5	0.007	0.0001 <sup>a</sup>	0.233
It was hard for me to walk one block when I had pain.	40.4	5	0.011	0.0001 <sup>a</sup>	0.450
It was hard to stay standing when I had pain.	58.6	5	0.004	0.0001 <sup>a</sup>	0.453
I had trouble sleeping when I had pain.	18.6	5	0.018	0.0023 <sup>a</sup>	0.175
It was hard for me to pay attention when I had pain.	11.5	5	0.021	0.0426	n.s.
<i>Anchor Item</i>					
<i>I felt angry when I had pain</i>					
<b>Fatigue SF</b>					
I got tired easily. <sup>b</sup>	25.2	5	0.025	0.0001	0.0698
I was too tired to enjoy the things I like to do.	n.s.				
I was too tired to do things outside.	n.s.				

Item: In the past 7 days...	$\chi^2$	df	Criterion $p^a$	$p$	wABC
I was so tired it was hard for me to pay attention.	n.s.				
Being tired made it hard for me to play or go out with my friends as much as I'd like.	n.s.				
I felt weak.	n.s.				
I had trouble starting things because I was too tired.	n.s.				
I had trouble finishing things because I was too tired.	n.s.				
Being tired made it hard for me to keep up with my schoolwork.	n.s.				
I was too tired to do sports or exercise.	n.s.				

<sup>a</sup>  $p$ -value based on Benjamini–Hochberg procedure.

<sup>b</sup> This was the only item with differential item functioning in step 1. df, degrees of freedom; wABC, expected score curves; SF, Short Form; n.s., non-significant.