



Published in final edited form as:

*Qual Life Res.* 2009 April ; 18(3): 359–370. doi:10.1007/s11136-009-9447-5.

## Measuring Global Physical Health in Children with Cerebral Palsy: Illustration of a Multidimensional Bi-factor Model and Computerized Adaptive Testing

**Stephen M. Haley, PhD, PT,**

Health and Disability Research Institute, Boston University Medical Campus, School of Public Health, 580 Harrison Ave, Boston MA 02218 (USA)

**Pengsheng Ni, MD, MPH,**

Health and Disability Research Institute, Boston University Medical Campus, School of Public Health, Boston MA

**Helene M. Dumas, MS, PT,**

Research Center for Children with Special Health Care Needs, Franciscan Hospital for Children, Boston, MA

**Maria A. Fragala-Pinkham, MS, PT,**

Research Center for Children with Special Health Care Needs, Franciscan Hospital for Children, Boston, MA

**Ronald K. Hambleton, PhD,**

Center for Educational Assessment, Department of Educational Policy, Research and Administration, University of Massachusetts, Amherst, MA

**Kathleen Montpetit, MScOT,**

Department of Occupational Therapy, Shriners Hospital for Children, Montreal

**Nathalie Bilodeau, MScOT,**

Department of Occupational Therapy, Shriners Hospital for Children, Montreal

**George E. Gorton, BS,**

Director, Motion Lab, Shriners Hospital for Children, Springfield, MA

**Kyle Watson, DPT, and**

Motion Analysis Lab, Shriners Hospital for Children, Philadelphia, PA

**Carole A Tucker, PT, PhD, PCS**

Associate Professor, College of Health Professions, Physical Therapy Department, Temple University, Philadelphia PA

### Abstract

**Purpose**—The purpose of this study was to apply a bi-factor model for the determination of test dimensionality and a multidimensional CAT using computer simulations of real data for the assessment of a new global physical health measure for children with cerebral palsy (CP).

**Methods**—Parent respondents of 306 children with cerebral palsy were recruited from four pediatric rehabilitation hospitals and outpatient clinics. We compared confirmatory factor analysis results across four models: (1) one-factor unidimensional; (2) two-factor multidimensional (MIRT); (3) bi-

factor MIRT with fixed slopes; and (4) bi-factor MIRT with varied slopes. We tested whether the general and content (fatigue and pain) person score estimates could discriminate across severity and types of CP, and whether score estimates from a simulated CAT were similar to estimates based on the total item bank, and whether they correlated as expected with external measures.

**Results**—Confirmatory factor analysis suggested separate pain and fatigue sub-factors; all 37 items were retained in the analyses. From the bi-factor MIRT model with fixed slopes, the full item bank scores discriminated across levels of severity and types of CP, and compared favorably to external instruments. CAT scores based on 10- and 15-item versions accurately captured the global physical health scores.

**Conclusions**—The bi-factor MIRT CAT application, especially the 10- and 15-item version, yielded accurate global physical health scores that discriminated across known severity groups and types of CP, and correlated as expected with concurrent measures. The CATs have potential for collecting complex data on the physical health of children with CP in an efficient manner.

### Keywords

Outcomes assessment; quality of life; item response theory; computerized adaptive testing; bi-factor model

---

### Introduction

Cerebral palsy (CP) is the most common physical disability in children and though non-progressive in nature, this disorder of movement and posture is characterized by physical impairments that impede function and health. Physical impairments most often associated with cerebral palsy include muscle weakness and spasticity, leading to joint stiffness and contractures. Children with CP also frequently have associated learning, vision, communication and seizure disorders. General physical health conditions such as pain [1–6] and fatigue [2,6,7] impact the quality of life of children with CP [2,4,8,9].

Using parents as proxies is a common method for obtaining information on health status and function for very young children and children with intellectual impairments.[10] Several parent-report health-related quality of life and functional outcome measures used for children with CP include questions about pain.[6,9,11,13] These instruments collectively contain a wide variety of pain items, yet each individual scale includes only a limited number of questions to assess the impact of pain on general physical health and well-being.

Also, children with CP often exhibit reduced muscular and cardiopulmonary endurance which can lead to fatigue [7]. For children with CP who are ambulatory, increased energy expenditure is needed for ambulation when compared to children without CP [14]. Most current health-related quality of life measures for children with CP do not include questions about the frequency, cause, or impact of fatigue on daily activities. The PedsQL[6] is a notable exception, but is limited to only four fatigue items.

Spasticity, an exaggeration of the tonic stretch reflex resulting in muscle hypertonicity, causes joint and body stiffness and is one of the characteristic physical impairments in CP. Medical, surgical, pharmaceutical and rehabilitation interventions are often directed at minimizing spasticity to prevent secondary conditions, improve the ease of care-giving and enhance functional performance. The amount of spasticity itself is often measured by clinicians but less effort has been directed toward obtaining parents' estimates of the impact of stiffness (resulting from spasticity) on daily routines.

Due to the multidimensionality of many health concepts, such as physical health, questionnaires have the potential to become very long and burdensome if all the dimensions are covered in one test battery [15]. To avoid being too burdensome, most surveys severely restrict the number of items to be administered. To circumvent questionnaires that are too extensive and unwieldy for children, computerized adaptive testing (CAT) is increasingly being proposed for use in routine functional and health related quality of life (HRQoL) assessments in pediatric rehabilitation [16] and appears to be the new trend of future patient-(parent and child) reported outcome assessments.[17–20]. CAT employs a simple form of artificial intelligence that selects questions tailored or matched to the patient, shortens or lengthens the test to achieve the desired level of score precision, and provides score estimates, regardless of the particular number of items administered, on a standard metric. Only enough items are administered in order to satisfy pre-set precision rules, or alternately, a maximum number of items are determined in advance. CAT platforms require the development of a comprehensive and calibrated set of items (item banks) that define each underlying dimension [21].

However, CAT applications require relatively strict adherence to the unidimensionality assumption of item response theory (IRT) analyses. IRT methods are used to estimate the position of items along the latent trait and then CAT software algorithms select items to match the child's ability level on that trait.[22] When IRT assumptions are met, score estimates do not strictly depend on a particular fixed set of items. This scaling feature allows one to compare children along a latent trait even if they have not completed the identical set of items. Since items and score estimates are defined on the same scale, items can be optimally selected to provide good score estimates of each level of the scale. This feature of IRT creates important flexibility in administering tests in a dynamic and tailored approach for each child.

The IRT unidimensionality assumption can be difficult to meet, as many of the quality of life dimensions are not strictly unidimensional. When a set of items includes more than one latent trait or dimension, we may consider multidimensional IRT (MIRT) as an option. There are two kinds of MIRT models. In the compensatory MIRT model, the deficiency in one trait can be offset by an increase in other abilities, whereas in the non-compensatory MIRT model, the deficiency in one trait can't be completely offset by an increase in others.[23] In MIRT models, we can often more fully understand the underlying structure behind the response pattern than with a single unidimensional model. In a CAT application, the MIRT model has been shown to be more efficient in estimating person scores.[24]

If data are truly multidimensional, then a bi-factor model using a (MIRT) model for CAT development is a potential approach to analyses and application. Bi-factor models allow each item to have a positive loading on a general factor that is thought to be the most important trait that underlies the item bank. Each item can then load on a specific content factor, from which sub-scores can be derived [25]. Reise [25] recently discussed both the merits and limitations of using a bi-factor model in health surveys. Under conditions in which two dimensions are at least moderately correlated, and items are loaded on both a general factor and specific content factors, a bi-factor model may be more appropriate than a unidimensional model, especially when factor loadings are higher on the sub-factor than the general factor [26]. However, the complexity of the MIRT model may limit its applicability and although it is robust, it needs adequate sample sizes to get fair parameter estimation.

Our first aim was to compare the confirmatory factor analysis results of unidimensional, two-factor and MIRT bi-factor models (varied and fixed discrimination) using items from a new global physical health scale developed for children with CP. The global physical health scale included fatigue, pain and joint stiffness items. Our second aim was to generate estimated general and content-specific scores from one of the models to determine whether the estimated child scores discriminated across levels of severity and types of CP, and how the scores

compared to external instruments. A third aim was to develop simulated CATs to determine the accuracy of CAT general factor and content sub-scores compared to those scores from the full item bank.

## Sample

We collected parent-report data on a convenience sample of 306 children with CP. Inclusion criteria were children with a known diagnosis of CP, ages 2–20 years, and parents with fluency in English. Also, children were excluded if they had received surgical or pharmacological interventions within the past six months. Data were collected across three Shriners Hospital for Children (SHC) orthopedic facilities in Philadelphia, Montreal and Springfield, MA and at Franciscan Hospital for Children (FHC), Boston, MA.

The mean age of the sample was 10.8 years (sd 4.0). Demographic characteristics of the sample are presented in Table 1. Our sample is not strictly representative of population data reported elsewhere, [27] as we have underrepresented children with more severe gross motor disabilities. Since two of the three SHC sites were motion lab based sites, it was easiest at those sites to recruit ambulatory participants; in some cases few non-ambulatory patients were seen. Human subject approval was obtained at each participating clinical institution and the Boston University Institutional Review Board.

## Global Physical Health Item Bank Construction

Following a thorough literature search, more than 50 adult and pediatrics measures of quality of life, function, general health, pain, fatigue and pain were compiled and reviewed to assess existing items in these areas to construct a comprehensive item bank. The construction of the item bank has been previously described, [28] The original global physical health item pool included 55 items that sampled the child's joint stiffness, pain, fatigue, and drooling. The final item pool, consisting of 37 items, was determined based on expert judgment of clinicians and researchers at SHC and FHC, results of cognitive testing [29], and an attempt to provide a comprehensive set of items with a consistent response scale that would span across different physical ability levels and ages.

## Data Collection Procedures

Global physical health items were administered to parents using a stand-alone PC-based tablet, in which parents reported on their child's physical health. Items were presented in a fixed order with no feedback between items. For a small number of parents (N=11), an internet version was made available because of the lack of feasibility of completing the survey in the clinic setting. Trained therapy and research staff were available to answer any questions regarding the study protocol or the interpretation of items, either in person or by phone. This study was part of a larger data collection effort in which item banks for three other scales (physical activity, lower extremity/movement skills and upper extremity skills) were also administered to parents at the same time. To minimize response fatigue for any one set of calibration items, the order of item banks was counterbalanced. Demographic information (ethnicity, sex, age, type of CP) was collected for each child. Severity of CP was initially rated by the parents and then confirmed by the research staff using the Gross Motor Function Classification System (GMFCS) [30], which rates children on a 5-point scale based primarily on ambulatory ability. In the GMFCS, the smaller number rating reflects better function. Type of CP (diplegia, hemiplegia and quadriplegia) was also initially rated by the parents and then confirmed by the research and therapy staff.

## Validity Measures

To serve as concurrent validity comparisons, The Pediatric Outcomes Data Collection Instrument (PODCI) [12] and the PedsQL™ Cerebral Palsy Module Version 3.0 (PedsQL-CP) [6] were also completed by parents. Within the calibration study, we recruited a convenience sample of 204 children whose parents completed the PODCI and 89 children whose parents completed the PedsQL-CP. The PODCI was developed specifically to assess changes following pediatric orthopedic interventions for a broad range of diagnoses, including children with CP. The PODCI is a standard outcome measure used routinely at the SHC. The PODCI covers a broad range of dimensions: upper extremity function, transfers and mobility, physical function and sports, comfort (lack of pain), happiness, satisfaction, and expectations. The most similar dimension to the new general health score and the content sub-scores was the PODCI comfort/pain subscale. This subscale includes three questions about amount of pain/discomfort and whether pain interferes with daily activities. The PedsQL-CP was designed to measure health related quality of life specifically in children with cerebral palsy. There are parent versions for toddlers (2–4 yrs), young children (5–7 yrs), child (8–12 yrs), and teens (13–18 yrs). All age versions were used in this study and combined. The domains include problems with: daily activities, school activities, movement and balance, pain and hurt, fatigue, eating, and speech and communication. Low scores indicate fewer problems and better quality of life. For comparison with the global physical health scores, we used the total PedsQL-CP (35 items), and the pain (4 items) and fatigue (4 items) sub-scores.

## Analyses

We conducted the analyses in three phases: (1) using confirmatory factor analysis, we compared the fit indices of a one-factor unidimensional, two-factor MIRT and two bi-factor MIRT models (fixed and varied slopes); (2) based on the fit indices and other considerations, we used one of the models to develop item parameters and a general factor and two content (fatigue and pain) person score estimates and tested whether the scores had discriminant and concurrent validity; and (3) conducted a series of computer simulation studies to examine the accuracy of CAT general factor and content sub-scores based to the full item bank scores. .

### Confirmatory factor analysis

The confirmatory factor analysis, which was based on maximum likelihood estimation, was used to help select the appropriate model. The model comparison was based on a likelihood ratio Chi Square, the number of different parameters estimated, and different information criteria indices: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and sample-size adjusted BIC, which takes into account the parsimony of the model. Models with lower information values indicate better fit, however more complex models with more parameters estimated lead to lower information values. See Kass and Wasserman for a full description of these indices [31]. We compared a one-factor model, a two-factor MIRT model, a bi-factor MIRT model with slope parameters fixed, and a MIRT bi-factor model with slope parameters allowed to vary. In the bi-factor model, the hypothesis is that there is a general factor which accounts for the commonality among the items, and multiple group factors which account for the unique influence of each content sub-domain above and beyond the general factor.[32] In a two-factor or multidimensional model, there is no general factor, items can load on one factor or cross-load on multiple factors.[33]

**Calculation of Person Score Estimates**—Items were calibrated based on the proportional odds logistic regression model, using Mplus software [34]. This model is similar to the Graded Response Model (GRM), as it estimates both item difficulty and discrimination parameters and orders each item category. The model belongs to the family of compensatory

models and it can be viewed as the extension of GRM, the category response probabilities can be represented as:

$$\begin{aligned} \text{First category: } p_{i1} &= \frac{1}{1 + \exp\left(\sum_{r=1}^R \lambda_{ir} f_r - \tau_{i1}\right)}; \\ \text{Middle category } p_{ij} &= \frac{1}{1 + \exp\left(\sum_{r=1}^R \lambda_{ir} f_r - \tau_{ij}\right)} - \frac{1}{1 + \exp\left(\sum_{r=1}^R \lambda_{ir} f_r - \tau_{ij-1}\right)}; \\ \text{Last category } p_{ij} &= 1 - \frac{1}{1 + \exp\left(\sum_{r=1}^R \lambda_{ir} f_r - \tau_{ij-1}\right)}; \end{aligned}$$

Where  $\lambda_{ir}$  is the factor loading for  $i^{th}$  item  $r^{th}$  factor,  $f_r$  is the factor score for  $r^{th}$  factor,  $\tau_{ij}$  is the threshold parameter for  $i^{th}$  item. We used this form of the GRM due to practical considerations of its polytomous extension availability within MPlus.

The item fit statistics were calculated based on the Z3 index, which is the standardized difference between the observed and expected log likelihood of response patterns. It is a one-sided test, and under the null hypothesis, the z scores are distributed as a normal distribution, and the cutoff is  $-1.65$  [35]. Originally the Z3 index was developed within a unidimensional item response model, but later extended to MIRT applications [36].

The general factor score was calculated using Bayesian estimation methods to maximize the variance for the general factor and two content sub-scores. In this situation, the subscale scores (pain and fatigue) reflect both the primary factor and the relevant secondary trait. To calculate the sub-scale scores, the relative weights of the factors contributing to an item response were determined and the subscale score was calculated as a weighted linear composite. The weights were chosen such that the sum of their squares was equal to 1 [37]. The first step was to determine the angle for each item relative to the general factor in each subtest

$\left(\alpha_i = \arccos \frac{\lambda_i}{\sqrt{\lambda_i^2 + \lambda_{i1}^2}}\right)$ , where  $\alpha_i$  is the angle for  $i^{th}$  item,  $\lambda_i$  is the general factor loading,  $\lambda_{i1}$  is the group factor loading. Secondly, we needed to get the average of the angle in each sub scale. Third, the composite scale was calculated as  $\theta_c = \cos(\bar{\alpha})\theta + \sin(\bar{\alpha})\theta_s$ , where  $\bar{\alpha}$  is the average angle,  $\theta$  and  $\theta_s$  is the general factor score and the group factor score. [37,38] Because group factor (pain and fatigue) was determined by the general factor, so the variation of the group factor was reflected by the variation of the general factor.

**Validity of Person Score Estimates**

We compared person score estimates from the general factor and the two content sub-scores of the global health scale to the PODCI comfort/pain subscale score and the PedsQL-CP pain and fatigue subscale scores using Pearson correlations. We predicted that general factor score would have a moderate to high correlation with the PedsQL-CP total score; the pain sub-score would have a moderate to high correlation with the PODCI comfort/pain and PedsQL-CP pain subscale scores, and the fatigue sub-score would have a moderate to high correlation with the PedsQL-CP fatigue score. We also examined the discriminant validity of the general factor score and content sub-score estimates on the severity of CP using the GMFCS severity levels and the type of CP. We hypothesized that children with more severe CP as indicated by GMFCS levels of IV and V (non-ambulatory levels) would have lower scores on the global health scales, and children with quadriplegia would have more physical health concerns than children with

either hemiplegia or diplegia. The ability of the global health scores to discriminate between groups of children based on levels of severity and type of CP was evaluated by comparing average group scores using one-way ANOVA tests with post-hoc (Tukey-Kramer test) comparisons.

### **CAT simulations**

We based the activity CAT algorithms on the HDRI™ software developed at the Health and Disability Research Institute. The score estimation in CAT was based on Bayesian modal estimation, which works to find the mode of the posterior distribution based on the Newton-Raphson procedure. In these simulations, responses to items selected by the CAT software were obtained for cases in the calibration data set and presented to the computer to simulate the conditions of an actual CAT assessment. The item selection rule was based on finding the item that maximized the determinant of the information matrix evaluated at the current score level. In the present study, we developed three CAT scores in the simulations to reflect three item-stopping rules based on number of items completed (CAT-15, CAT-10 and CAT-5) [39]. These simulated scores were compared to the latent trait global physical health scores estimated by the full item bank. We also compared the discriminant validity of the scores from the different CAT versions including the scores developed from the full item bank.

## **Results**

### **Confirmatory factor analysis**

We examined four models that retained all 37 global physical health items. We compared the estimated IRT item discrimination parameters for a one-factor unidimensional model, a two-factor MIRT model with each item loading on only one factor, and two MIRT bi-factor models. We calculated a likelihood ratio Chi Square and three information criteria (AIC, BIC, and sample-size adjusted BIC). In Table 2, note the differences in the log-likelihood values, information criteria and in the number of parameters that are needed to be estimated. Using the differences in the log-likelihood values, we found the two-factor model to have significantly better fit than the one-factor model (Chi Square = 602.59,  $p < 0.001$ ), while the varied slope bi-factor model fit better than the fixed bi-factor model (Chi Square 618.58;  $p < 0.001$ ). There is general favor for choosing model solutions that have higher values of the observations per parameter ratio.[40] Given our current sample size, the fact that the observations per parameter ratio in the bi-factor model with fixed slope is 2.1 as compared to 1.4 in the bi-factor model with varied slopes, we chose to do further analyses with the fixed slope MIRT bi-factor model.

### **Validity of General Factor and Content Sub-scores**

All 37 items were retained in the final item bank. Seven items loaded on the general factor only, while 17 items loaded both on general and fatigue sub-factor, and 13 loaded both on the general and pain sub-factor. Based on the  $Z^3$  criteria of  $< -1.65$  for item fit, two items (“how often is your child tired when doing homework?”, “how much pain does your child have when wearing braces?”) misfit the model. We chose not to remove those items because of the importance of the content (e.g., over 75% of the sample wore braces). All items with pain content loaded on the general and pain factors. All items with fatigue content loaded on the general and fatigue factors. The rest of the items loaded only on the general factor. See Table 3 for a listing of item loadings, item parameters and item fit. The correlation of estimated person scores for the general factor and pain factor was 0.81, and the correlation between the general factor and fatigue factor was 0.95; the correlation between the pain and fatigue sub-factors was 0.70. These high correlations were expected due to the fact that each of the content sub-scores were determined by information from both the general factor items and items from each sub-domain.

We compared the general factor and content sub-scores from the bi-factor MIRT model with fixed slopes to two discriminant criteria: the GMFCS severity levels, and type of CP. The general factor scores were able to discriminate across GMFCS levels ( $F(3,299) = 32.82$ ,  $p < 0.001$ ); post-hoc analyses revealed differences between IV/V and all other levels, and between levels I and II, and I and III. The fatigue sub-score was also able to discriminate across GMFCS levels ( $F(3,299) = 40.85$ ,  $p < 0.001$ ); with the post-hoc results the same as with the general factor score. The pain sub-score could also discriminate across GMFCS levels ( $F(3,299) = 19.13$ ,  $p < 0.001$ ); post-hoc analyses revealed that differences could be detected between Levels IV/V and the Levels I, II, and III. These results are summarized in Figure 1. We found differences in the general factor scores across types of CP, ( $F(2,300) = 33.78$ ;  $p < 0.0001$ ). Post-hoc analyses indicated that the general factor score could discriminate across children with quadriplegia, hemiplegia and diplegia. The fatigue sub-score discriminated across CP types ( $F(2,300) = 46.24$ ;  $p < 0.0001$ ); significant difference were noted between children with hemiplegia, diplegia and quadriplegia, with children with hemiplegia having the least problems with fatigue. The pain scale differentiated among CP types ( $F(2,300) = 16.86$ ;  $p < 0.0001$ ); post-hoc analyses indicated that the pain sub-score could discriminate between children with quadriplegia versus children with either hemiplegia or diplegia. See Figure 2.

We also compared the general factor and content sub-scores from the bi-factor MIRT model with fixed slopes to two concurrent measures, namely the PODCI and PedsQL-CP. The general factor score ( $r = 0.54$  and the fatigue score ( $r = 0.46$ ) were correlated moderately with summary scores from the PODCI comfort (pain) scale. As expected, the pain sub-score was correlated with the PODCI comfort (pain) scale ( $r = 0.63$ ). The general factor score was correlated with the PedsQL-CP total score ( $r = -0.75$ ). The fatigue sub-score was correlated with the PedsQL-CP fatigue items ( $r = -0.68$ ) and the pain sub-score was correlated with the PedsQL-CP pain items ( $r = -0.65$ ). The negative correlations reflect that the higher scores on the PedsQL-CP indicate more problems, while the higher scores on the global physical health scores indicate higher levels of physical health. A full correlation matrix is presented in Table 4.

### CAT simulations

As reported in Table 5, the scores from the 10- and 15-item simulation CAT were quite similar to those for the full item bank score. Even the 5-item CAT scores had fairly high correlations ( $r = 0.885 - 0.912$ ) with the item bank scores. Note though that these results are somewhat biased because of the inclusion of items used in the CATs come directly from in the item bank as well. Normally the bias would be quite small in these types of studies, but in this study, the item bank itself had only 37 items. We suspect, as in other studies, a CAT of 10 to 15 items would be necessary to replicate scores from a much larger item bank than was used in this study. We also found that the 10- and 15-item CAT versions had only slightly decreased discriminatory power as compared to the full set of items to differentiate across CP gross motor severity levels and across types of CP. The discrimination lost appreciable accuracy in the 5-item CATs. See Table 6.

### Discussion

The assessment of general physical health, including components of pain and fatigue, are key components of an overall health and functional evaluation for children with CP [41,42] undergoing surgical or other interventions. The results of this study suggest that a global physical health scale with three generated person scores might provide a uniform assessment approach for children with CP across a wide age range, types and across multiple levels of severity. An advantage of the CAT format over more traditional fixed forms is that more content becomes available to a child in the new measure (17 fatigue items, 13 pain items) than in most Quality of life measures that are limited to 3–5 items per scale.



The bi-factor model was useful in this study, as we combined similar, but apparently distinct constructs of global physical health in one set of items. Based on the confirmatory factor analysis results, we found that this set of items had better fit with a bi-factor model than a more commonly used unidimensional model, and had better fit than a two-factor model. Even when non-pain and non-fatigue items were removed from the 2-factor solution, the bi-factor model with fixed slope still had better fit than the 2-factor model. In this situation, most of the items were correlated with the general factor and one of two content factors (pain and fatigue). For the pain items, the correlation was greater within the pain scale than the general factor, indicating the appropriateness of the bi-factor model. The potential advantage of using the bi-factor MIRT model was to not only generate general factor scores, but also to create sub-scores on the two content domains. Use of the bi-factor MIRT model has not been commonly addressed in health care applications, but can have immediate clinical appeal.

In our results to date, there may be considerable redundancy between the general factor score and the fatigue sub-score ( $r=0.95$ ). This is also evident in the patterns of score discrimination with levels of severity of CP and types of CP (Figures 1 and 2). We have chosen to keep these two scores in the model, since the fatigue factor is assuming additional variance beyond that of the general factor, leading to better overall model fit (Table 3). It is not surprising that general and content scores are correlated as information from the general scale is used partially for determination of the content scores.

We found the general factor and fatigue scores to discriminate among levels of gross motor severity as expected, with children with more severe types of CP who were non-ambulatory having more general health concerns. The general factor score could not discriminate between severity Levels II (walks with limitations) and III (walks using a hand-held mobility device). In contrast, the pain sub-factor scores discriminated between the children with the most severe CP (Levels IV- self-mobility with limitations; may use powered mobility and Level V- transported in a manual wheelchair) and children at each of the other levels in which some type of ambulation occurred. We can speculate that greater pain in these children categorized in Levels IV and V may be due either to prolonged sitting positions, inability to weight shift and change position, or increased prevalence of orthopedic complications such as fractures and contractures.

The concurrent validity results help confirm the usefulness of the specific content sub-scores. We found a strong correlation with the PODCI pain sub-score ( $r = 0.71$ ). Similarly, the pain and fatigue sub-scores on the PedsQL-CP showed good correspondence to the pain and fatigue sub-scores of the global health measure. (Table 4). These results suggest that the pain and fatigue sub-scale are useful indices to report in addition to the general factor score

The results of the CAT simulations imply that 10, and 15 item versions yield reasonably accurate estimates of global physical health in children with CP. The 15-item CAT has excellent reproducibility with the full item bank. We note some decrement in accuracy in the five-item CAT, particularly with the pain and fatigue scores. We especially want to discourage the use of the five-item CATs based upon the findings from this study since the results we have reported probably have a modest positive bias. These overall results are quite comparable to several other pediatric studies that have conducted simulations using real datasets, comparing responses to all items in the item bank to a CAT simulation of various lengths [24,43–45]. Computer simulations are conducted under the assumption that answers to a subset of items selected using CAT would be identical to the answers given when the items were embedded in the full-item parent report version. They also assume the item response model is correct and is the basis for item responses. On the other hand, a limitation of using a real data set for the simulation study is that it is difficult to know how the results will generalize to other situations and we assume that context does not impact on item responses [37].

These preliminary data indicate that the 15-item CAT version has very good discrimination abilities, approaching the level of the full item bank. This discrimination ability is slightly diminished with the 10-item CAT, and is considerably reduced with the 5-item CAT. . Apparently, these relationships become less consistent as the number of items used to estimate scores is reduced. Future studies on the responsiveness of this global health scale for children with CP are planned. If successful, CAT versions will be used to evaluate global health changes in children with CP after orthopedic surgeries and conservative interventions such as bracing, therapy, and spasticity management (medications and botulinum toxin or phenol injections) within the SHC system.

There are a number of study limitations that we want to note. The sample size is somewhat marginal for sufficiently stable and reliable parameter estimates, especially with such rather complex models employing slope parameters and parameters for the general factor and content sub-factors. We did try to minimize effects of the small sample size by fixing the slope of the bi-factor model to reduce the number of parameters that needed to be estimated. We also point out that the sample is somewhat non-representative of children with CP, as more ambulatory children are represented in the sample than what is normally seen in the population. This was due, in part, to the emphasis of the gait labs within the SHC, and of the typical children with CP that are routinely part of their systems. At this stage of our calibration work, we have used the analysis of a real data set with its *a-priori* unknown structure for estimates of how the CAT programs will work in practice. These simulation studies with real calibration data can provide overestimates of how CATs perform in prospective studies. And finally, as a first step, we chose to use parent ratings only; in the future we would like to extend this work to offer a child-report version as well.

We believe that the results of this study suggest that a single instrument with a bi-factor MIRT model and its CAT applications can generate meaningful general factor and content domain sub-scores to assess the general physical health of children with CP. Although the bi-factor MIRT model may be conceptually appealing, it is a complex model, and it will be sometime before these applications will become routine. We submit that these data support the further development and testing of bi-factor models for future CAT development. For certain, we would like to see future studies carried out with larger sample sizes so that the model item parameters can be estimated with more precision.

## Acknowledgments

Supported by the Shriners Hospital for Children Foundation (Grant # 8957) and an Independent Scientist award to Dr. Haley (National Center on Medical Rehabilitation Research/NICHD/NIH, grant # K02 HD45354-01A1).

## References

1. Engel J, Petrina T, Dudgeon B, McKearnan K. Cerebral palsy and chronic pain: a descriptive study of children and adolescents. *Physical & Occupational Therapy in Pediatrics* 2006;25:73–84. [PubMed: 16418117]
2. Berrin S, Malcarne V, Varni J, Burwinkle T, Sherman S, Artavia K, et al. Pain, fatigue, and school functioning in children with cerebral palsy: a path-analytic model. *Journal of Pediatric Psychology* 2007;32:330–337. [PubMed: 16818483]
3. Tervo R, Symons F, Stout J, Novacheck T. Parental report of pain and associated limitations in ambulatory children with cerebral palsy. *Archives of Physical Medicine and Rehabilitation* 2006;87:928–934. [PubMed: 16813780]
4. Houlihan C, O'Donnell M, Conaway M, Stevenson R. Bodily pain and health-related quality of life in children with cerebral palsy. *Developmental Medicine and Child Neurology* 2004;46:305–310. [PubMed: 15132260]

5. Castle K, Imms C, Jun H. Being in pain: a phenomenological study of young people with cerebral palsy. *Developmental Medicine & Child Neurology* 2007;49:445–449.
6. Varni J, Burwinkle T, Berrin S, Sherman S, Artavia K, Malcarne V, et al. The PedsQL in pediatric cerebral palsy: reliability, validity, and sensitivity of the Generic Core Scales and Cerebral Palsy Module. *Developmental Medicine and Child Neurology* 2006;48:442–449. [PubMed: 16700934]
7. Rimmer J. Physical fitness levels of persons with cerebral palsy. *Developmental Medicine and Child Neurology* 2001;43:208–212. [PubMed: 11263693]
8. Vargus-Adams J. Longitudinal use of the Child Health Questionnaire in childhood cerebral palsy. *Developmental Medicine and Child Neurology* 2006;48:343–347. [PubMed: 16608541]
9. Narayanan U, Fehlings D, Weir S, Knights S, Kiran S. Initial development and validation of the Caregiver Priorities and Child Health Index of Life with Disabilities (CPCHILD). *Developmental Medicine & Child Neurology* 2006;48:804–812. [PubMed: 16978459]
10. White-Koning M, Arnaud C, Bourdet-Lubere S, Bazex H, Colver AF, Grandjean H. Subjective quality of life in children with intellectual impairment--how can it be assessed? *Developmental Medicine and Child Neurology* 2005;47:281–285. [PubMed: 15832552]
11. Novacheck T, Stout J, Tervo R. Reliability and validity of the Gillette Functional Assessment Questionnaire as an outcome measure in children with walking disabilities. *Journal of Pediatric Orthopedics* 2000;20:75–81. [PubMed: 10641694]
12. Daltroy L, Liang M, Fossel A, Goldberg M. The POSNA Pediatric Musculoskeletal Functional Health Questionnaire: report on reliability, validity, and sensitivity to change. *Journal of Pediatric Orthopedics* 1998;18:561–571. [PubMed: 9746401]
13. Landgraf, J.; Abetz, L.; Ware, JE, Jr. *The CHQ: A User's Manual*. Vol. 1. Boston, MA: The Health Institute, New England Medical Center; 1996.
14. Bottos M, Gericke C. Ambulatory capacity in cerebral palsy: prognostic criteria and consequences for intervention. *Developmental Medicine & Child Neurology* 2003;45:786–790. [PubMed: 14580136]
15. Ware JE Jr. Conceptualization and measurement of health-related quality of life: comments on an evolving field. *Archives of Physical Medicine and Rehabilitation* 2003;84:S43–51. [PubMed: 12692771]
16. Jacobusse G, van Buuren S. Computerized adaptive testing for measuring development of young children. *Statistics in Medicine* 2007;26:2629–2638. [PubMed: 17133649]
17. Revicki D, Cella D. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Quality of Life Research* 1997;6:595–600. [PubMed: 9330558]
18. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short forms, and computerized adaptive assessment. *Quality of Life Research* 2007;16:133–141. [PubMed: 17401637]
19. Hays R, Lipscomb J. Next steps for use of item response theory in the assessment of health outcomes. *Quality of Life Research* 16(Suppl 1):195–199.
20. Ware JE Jr, Gandek B, Sinclair S, Bjorner B. Item response theory in computer adaptive testing: implications for outcomes measurement in rehabilitation. *Rehabilitation Psychology* 2005;50:71–78.
21. Wainer, H. *Computerized Adaptive Testing: A Primer*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
22. Hambleton, RK. *Educational Measurement*. Vol. 3. New York: American Council on Education-Macmillan Publishing Company; 1989. Principles and selected applications of Item Response Theory; p. 147-200.
23. Ackerman T. Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement* 1989;13:113–127.
24. Haley S, Ni P, Ludlow L, Fragala-Pinkham M. Measurement precision and efficiency of multidimensional computer adaptive testing of physical functioning using the Pediatric Evaluation of Disability Inventory. *Archives of Physical Medicine and Rehabilitation* 2006;87:1223–1229. [PubMed: 16935059]
25. Reise S, Morizot J, Hays R. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research* 2007;16(Suppl 1):19–31. [PubMed: 17479357]

26. Reeve B, Hays R, Bjorner J, Cook K, Crane P, Teresi J, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care* 2007;45(5 Suppl 1):S22–31. [PubMed: 17443115]
27. Gorter J, Rosenbaum P, Hanna S, Palisano R, Bartlett D, Russell D, et al. Limb distribution, motor impairment, and functional classification of cerebral palsy. *Developmental Medicine & Child Neurology* 2004;46:461–467. [PubMed: 15230459]
28. Tucker C, Haley S, Watson K, Dumas H, Fragala-Pinkham M, Gorton G, et al. Physical function for children and youth with cerebral palsy: Item bank development for computer adaptive testing. *Journal of Pediatric Rehabilitation Medicine* 2008;1:237–244. [PubMed: 19779597]
29. Dumas H, Watson K, Fragala-Pinkham M, Haley S, Bilodeau N, Montpetit K, et al. Cognitive interviewing to elicit parent feedback of test items for assessing physical function in children with cerebral palsy. *Pediatric Physical Therapy* 2008;20:356–362. [PubMed: 19011526]
30. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine and Child Neurology* 1997;39:214–223. [PubMed: 9183258]
31. Kass R, Wasserman L. A reference Bayesian test for nested hypotheses and its relationship to the Schwartz criterion. *Journal of the American Statistical Association* 1995;90:928–934.
32. Chen F, West S, Sousa K. A comparison of bifactor and second-order models of quality of life. *Multivariable Behavioral Research* 2006;41:189–225.
33. Adams R, Wilson M, Wang W. The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement* 1997;21:1–23.
34. Muthen, B.; Muthen, L. *Mplus User's Guide*. Los Angeles: Muthen & Muthen; 2001.
35. Dragow F, Levine M, Williams E. Appropriateness measurement with polytomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology* 1985;38:67–86.
36. Ackerman, T.; Hombo, C.; Neustel, S. Evaluating indices used to assess the goodness-of-fit of the compensatory multidimensional item response theory model. Poster presented at the Annual Meeting of the National Council on Measurement in Education; New Orleans LA. 2002.
37. DeMars, C. Scoring subscales using multidimensional item response theory models. Poster presented at the annual meeting of the American Psychological Association; Washington, DC. 2005.
38. Ackerman T. Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education* 1994;7:255–278.
39. Segall D. Multidimensional adaptive testing. *Psychometrika* 1996;61:331–354.
40. Jackson D. Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling* 2003;1:128–141.
41. Schenker R, Coster W, Parush S. Participation and activity performance of students with cerebral palsy within the school environment. *Disability and Rehabilitation* 2005;27:539–552. [PubMed: 16019863]
42. Harvey A, Robin J, Morris M, Graham H, Baker R. A systematic review of measures of activity limitation for children with cerebral palsy. *Developmental Medicine & Child Neurology* 2008;50:190–198. [PubMed: 18201218]
43. Coster W, Haley S, Ni P, Dumas H, Fragala Pinkham M. Assessing self-care and social function using a computer adaptive testing version of the pediatric evaluation of disability inventory. *Archives of Physical Medicine & Rehabilitation* 2008;89:622–629. [PubMed: 18373991]
44. Haley S, Raczek A, Coster W, Dumas H, Fragala-Pinkham M. Assessing mobility in children using a computer adaptive testing version of the Pediatric Evaluation of Disability Inventory (PEDI). *Archives of Physical Medicine & Rehabilitation* 2005;86:932–939. [PubMed: 15895339]
45. Haley S, Ni P, Fragala-Pinkham M, Skrinar A, Corzo D. A computer adaptive testing approach for assessing physical function in children and adolescents. *Developmental Medicine and Child Neurology* 2005;47:113–120. [PubMed: 15707234]

## Appendix: Global Physical Health Scale

How often is your child's neck stiff?

- How often is your child's body stiff?
- How often is at least one of your child's arms stiff?
- How often is at least one of your child's legs stiff?
- How often does your child have trouble holding his/her head up when he/she is physically tired?
- How often does your child get physically tired when sitting in a chair or wheelchair at home or at school for more than one hour?
- How often does your child have trouble changing positions because he/she is physically tired?
- How often does your child have trouble with transfers because he/she is physically tired?
- How often does your child have trouble standing for 5 minutes because he/she is physically tired?
- How often does your child have trouble walking because he/she is physically tired?
- How often does your child have trouble climbing a flight of 8–12 steps because he/she is physically tired?
- How often does your child have trouble running because he/she is physically tired?
- How often does your child have trouble playing games or sports with other children because he/she is physically tired?
- How often does your child trip and fall because he/she is physically tired?
- How often does your child have difficulty getting around by him/herself because he/she is physically tired?
- How often does your child have difficulty with school activities because he/she is physically tired?
- How often does your child have difficulty concentrating on homework, reading, or quiet activities because he/she is physically tired?
- How often does your child need time during the day to rest?
- How often does your child have difficulty doing activities after a day of school because he/she is physically tired?
- How often does your child miss school because he/she is physically tired?
- How often does physical pain make falling asleep at night hard for your child?
- How often does physical pain wake up your child at night?
- How often does your child have physical pain when sitting in a chair or wheelchair?
- How often does your child have physical pain with transfers?
- How often does your child have physical pain when changing positions?

How often does your child have physical pain when standing?

How often does your child have physical pain when moving around?

How often does your child have physical pain when climbing one flight of 8–12 steps?

How often does your child have physical pain when running?

For this question, “playing games and sports” means activities that children play with other children such as tag, basketball or bowling.

How often does your child have physical pain when playing games or sports?

How often does your child have physical pain when riding in a car, bus or train?

How often does your child miss school because of physical pain?

How often does your child participate in indoor recreational activities without getting physically tired?

How often is your child physically active?

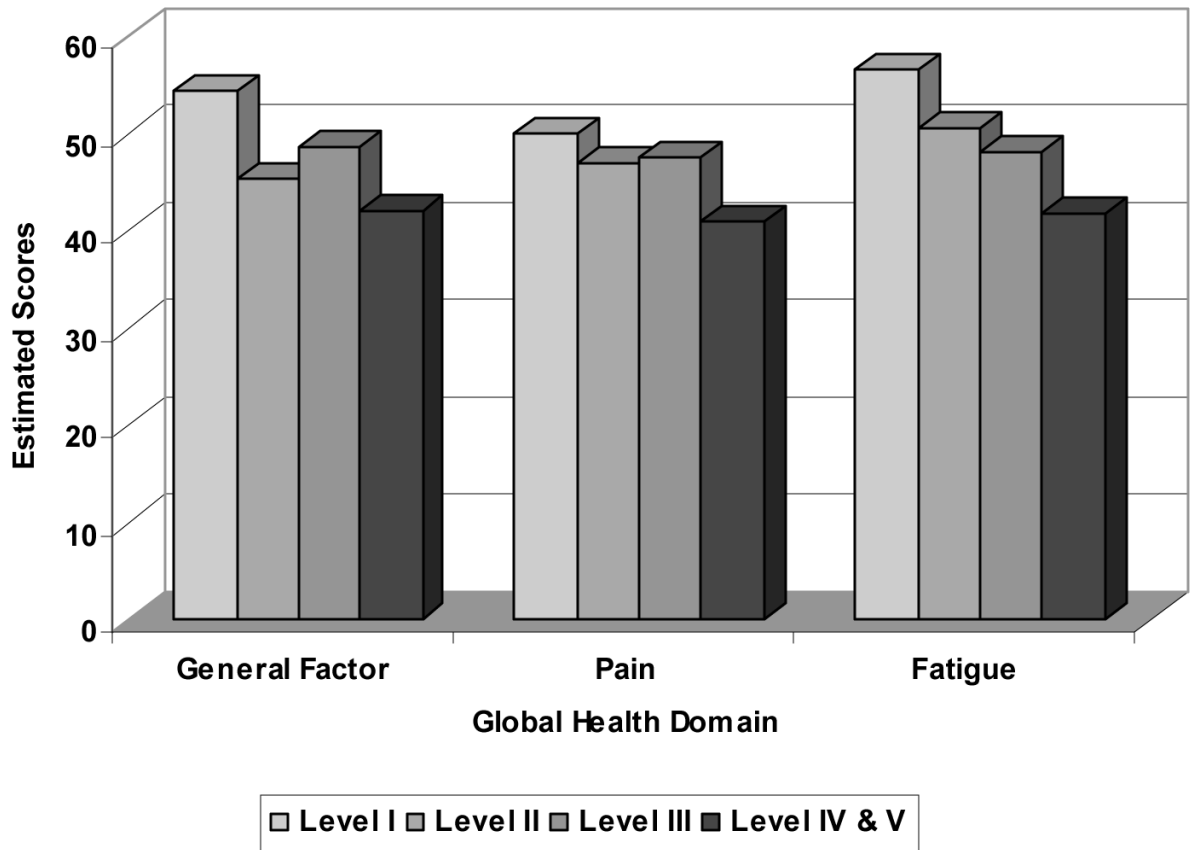
How often does your child’s physical pain interfere with wearing his/her braces or splints?

How often does your child’s physical pain interfere with using his/her adaptive equipment?

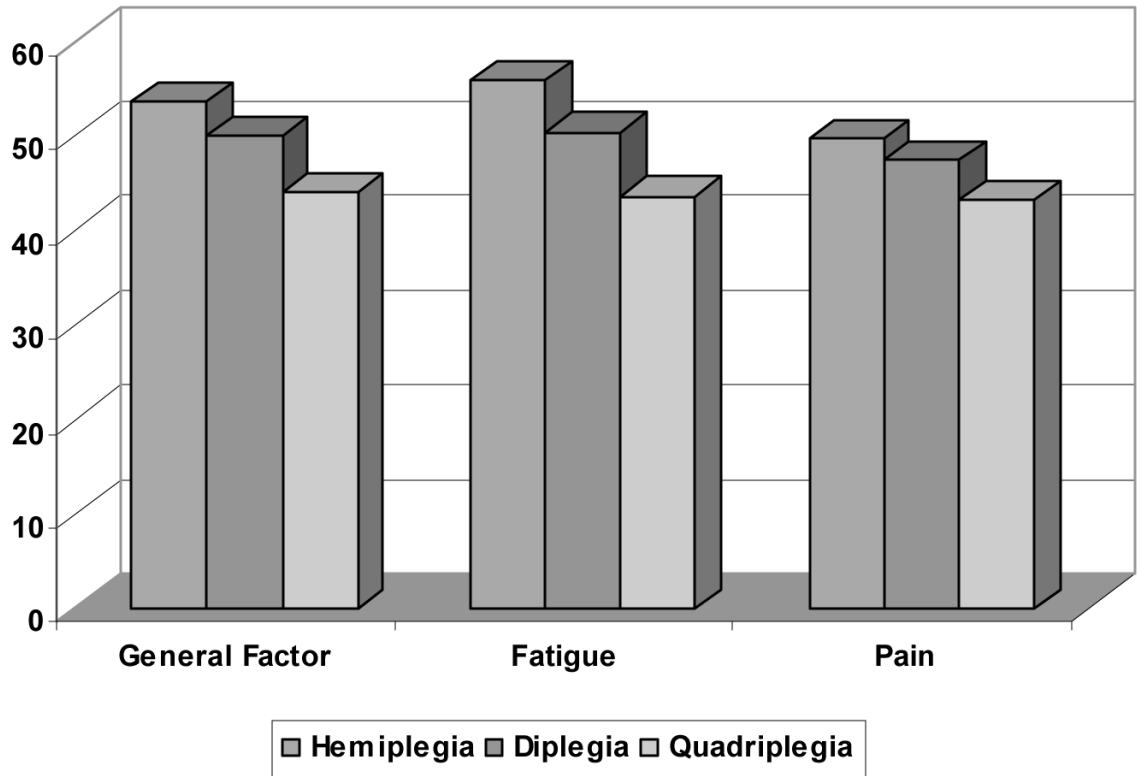
How often does your child drool when sitting quietly?

How often does your child drool with activity?

Rating scale: Always; Often; Sometimes; Rarely; Never; Not Applicable



**Figure 1.** Discriminant Validity of Global Physical Health- Average General Factor Score, Pain and Fatigue Sub-scores by Gross Motor Severity



**Figure 2.** Discriminant Validity of Global Physical Health- Average General Factor Score, Pain and Fatigue Sub-scores by Type of CP



**Table 1**

## Demographics

<b>Age Range</b>	2–20 years
<b>Age (mean) (SD)</b>	10.8 (4.0)
<b>Age Groups N (%)</b>	
< 5 yrs	17 (5.6)
5–9	112 (36.6)
10–14	123 (40.2)
15–19	50 (16.3)
≥20	4 (1.3)
<b>Female N (%)</b>	138 (45.1)
<b>Ethnicity N (%)</b>	25 (8.2)
Hispanic or Latino	
<b>Race N (%)</b>	9 (3.0)
Asian	
Other	12 (3.9)
African American	27 (8.8)
White	258 (84.3)
<b>Gross Motor Functional Classification System (GMFCS) N (%)</b>	
I	75 (24.5)
II	91 (29.7)
III	78 (24.5)
IV	36 (11.8)
V	26 (8.5)
<b>Type of CP N (%) (N=305)</b>	
Diplegia	144 (47.2)
Hemiplegia	73 (23.9)
Quadriplegia	88 (28.9)
<b>Environment/School count N (%) (N=305)</b>	
No school	13 (4.3)
Pre-school	11 (3.6)
Day school	217 (71.1)
Residential school	45 (14.8)
Home school	13 (4.3)
College/university, lives at home	4 (1.3)
College/university, lives on campus	2(0.7)
<b>Mode of administration N (%)</b>	
Internet	11(3.6)
In clinic	295 (96.4)
<b>Respondents education level N (%) (N=246)</b>	
Less than 9 <sup>th</sup> grade	2(0.8)
Some high school	11(4.5)
High school graduate	58(23.6)
Some college	60(24.4)
College graduate	79(32.1)

Graduate school

36(14.6)

---

Values are based on N=306 unless otherwise specified.

**Table 2**  
Model Comparisons Based on Log-likelihood and Fit Indices

	Unidimension Model	Multidimension Model	Bi-factor Varied Slope Model	Bi-factor Fixed Slope Model
Log-likelihood	-10777.303	-10476.007	-10188.801	-10498.094
Number of Free Parameters	182	183	219	148
AIC	21918.605	21318.015	20815.603	21292.188
BIC	22596.297	21999.431	21631.068	21843.279
Sample-Size Adjusted BIC	22019.077	21419.039	20936.501	21373.891

**Table 3**  
Item Parameters, Item Fit and Fixed Discrimination Results of MIRT Bi-factor Model

	General Factor	Sub-factor-Pain	Sub-factor - Fatigue	Step1	Step2	Step3	Step4	Item Fit*
neck stiff	1.615	0	0	-4.766	-4.078	-2.293	-0.225	1.98
body stiff	1.615	0	0	-2.85	-1.493	-0.189	1.517	3.91
arm stiff	1.615	0	0	-2.459	-1.518	-0.674	0.63	2.24
leg stiff	1.615	0	0	-1.343	-0.057	1.538	3.106	0.8
tired head up	1.615	0	1.117	-4.291	-2.795	-1.621	-0.245	2.2
tired sit	1.615	0	1.117	-5.264	-3.557	-1.404	0.343	-1.33
tired change position	1.615	0	1.117	-3.526	-2.868	-1.407	0.06	2.13
tired transfer	1.615	0	1.117	-3.32	-2.787	-1.092	0.584	3.67
tired stand	1.615	0	1.117	-3.496	-2.135	-0.28	1.441	2.05
tired walk	1.615	0	1.117	-3.766	-1.511	0.971	3.334	2.88
tired stairs	1.615	0	1.117	-3.476	-1.749	0.454	2.686	2.78
tired run	1.615	0	1.117	-3.328	-1.361	1.169	3.308	2.69
tired sports	1.615	0	1.117	-2.843	-1.733	0.508	2.292	1.5
tired fall	1.615	0	1.117	-4.28	-1.616	0.926	3.124	-1.31
tired get around	1.615	0	1.117	-5.98	-3.299	-0.399	2.002	3.54
tired at school	1.615	0	1.117	-5.68	-3.568	-0.694	1.729	3.16
<i>tired homework</i>	<b>1.615</b>	<b>0</b>	<b>1.117</b>	<b>-4.804</b>	<b>-2.042</b>	<b>0.013</b>	<b>1.892</b>	<b>-2.56*</b>
rest	1.615	0	0	-4.029	-2.916	-0.867	1.581	-0.9
tired after school	1.615	0	1.117	-5.477	-2.786	-0.157	2.118	1.24
tired missing school	1.615	0	1.117	-7.776	-5.517	-3.694	-1.665	0.15
pain sleep	1.615	1.923	0		-5.261	-2.925	-0.564	3.11
pain wake up	1.615	1.923	0		-5.927	-3.204	-0.726	2.32
pain sit	1.615	1.923	0	-7.055	-6.104	-3.63	-1.025	4.32
pain transfer	1.615	1.923	0	-6.947	-5.65	-3.716	-1.245	3.29
pain change position	1.615	1.923	0	-7.4	-6.247	-3.686	-1.212	3.79
pain stand	1.615	1.923	0	-6.226	-4.182	-1.72	1.04	4.28
pain move	1.615	1.923	0	-7.23	-5.209	-2.567	0.245	2.9
pain stairs	1.615	1.923	0	-6.108	-4.156	-2.071	0.28	4.97
pain running	1.615	1.923	0	-5.193	-3.741	-1.608	0.989	3.96
pain play	1.615	1.923	0	-6.278	-4.398	-2.002	0.772	2.19

	General Factor	Sub-factor-Pain	Sub-factor - Fatigue	Step1	Step2	Step3	Step4	Item Fit*
pain riding	1.615	1.923	0	-7.815	-6.588	-4.188	-0.817	2.43
pain missing school	1.615	1.923	0		-7.201	-4.934	-1.967	0.47
tired indoor activity	1.615	0	1.117	-6.404	-4.856	-2.341	0.122	0.3
tired activity	1.615	0	1.117	-5.237	-3.256	-0.775	1.685	-0.28
<i>pain wearing braces</i>	<b>1.615</b>	<b>1.923</b>	<b>0</b>	<b>-6.795</b>	<b>-4.447</b>	<b>-1.624</b>	<b>0.889</b>	<b>-1.85*</b>
drool sit	1.615	0	0	-3.971	-2.881	-1.793	-0.862	2.35
drool activity	1.615	0	0	-4.251	-2.797	-1.839	-1	2.45

\* <-1.65 misfit

**Table 4**  
 Concurrent Validity (PedsQL-CP and PODCI vs Global Physical Health Summary Scores)

	Global Score	Pain	Fatigue
<b>PedsQL-CP</b>			
Overall	-0.7545	-0.5697	-0.7124
Daily Activity	-0.6271	-0.3742	-0.5797
School Activity	-0.6806	-0.5418	-0.6454
Move/Balance	-0.6881	-0.5102	-0.6458
Pain/Hurt	-0.4744	-0.6524	-0.4071
Fatigue	-0.6637	-0.5338	-0.6836
Eating	-0.6140	-0.4601	-0.5863
Speech and Communication	-0.3169	-0.2267	-0.3078
<b>PODCI</b>			
Pain/Comfort	0.5540	0.7087	0.4612
Upper Extremity	0.5174	0.3551	0.5406
Transfers/Basic Mobility	0.5860	0.4554	0.6122
Sports/Physical Function	0.6609	0.4925	0.7019
Expectations	-0.0843	-0.0846	-0.1196
Happiness	0.4373	0.4374	0.4147
Happy/Satisfied	0.4520	0.4437	0.4327
Global Function	0.7064	0.5746	0.7372

**Table 5**

Pearson Correlations between Global Physical Health Estimated Scores Generated by CAT Program and Scores from Full Item Bank

	General Factor Score	Fatigue Sub-score	Pain Sub-score
CAT-15	0.960	0.957	0.976
CAT-10	0.933	0.929	0.963
CAT-5	0.893	0.885	0.912

**Table 6** Comparisons of Validity Tests using Relative Efficiency between Scores from Full item Bank and CATs

	Full Item Bank	CAT-15	CAT-10	CAT-5
Discriminant Validity with GMFCS	. <i>General factor</i> score	0.960	0.874	0.697
	<i>Fatigue</i> sub-factor score	1.000	0.996	0.871
	<i>Pain</i> sub-factor score	0.983	0.842	0.693
Discriminant Validity with type of CP	<i>General factor</i> score	0.946	0.851	0.643
	<i>Fatigue</i> sub-factor score	1.000	0.942	0.718
	<i>Pain</i> sub-factor score	0.934	0.782	0.730