Factorial invariance of pediatric patient self-reported fatigue across age and gender: a multigroup confirmatory factor analysis approach utilizing the PedsQL^TM Multidimensional Fatigue Scale

**James W. Varni · A. Alexander Beaujean ·
Christine A. Limbers**

Abstract

Objectives In order to compare multidimensional fatigue research findings across age and gender subpopulations, it is important to demonstrate measurement invariance, that is, that the items from an instrument have equivalent meaning across the groups studied. This study examined the factorial invariance of the 18-item PedsQL^TM Multidimensional Fatigue Scale items across age and gender and tested a bifactor model.

Methods Multigroup confirmatory factor analysis (MG-CFA) was performed specifying a three-factor model across three age groups (5–7, 8–12, and 13–18 years) and gender. MG-CFA models were proposed in order to compare the factor structure, metric, scalar, and error variance across age groups and gender. The analyses were based on 837 children and adolescents recruited from general pediatric clinics, subspecialty clinics, and hospitals in which children were being seen for well-child checks, mild acute illness, or chronic illness care.

Results A bifactor model of the items with one general factor influencing all the items and three domain-specific factors representing the General, Sleep/Rest, and Cognitive Fatigue domains fit the data better than oblique factor models. Based on the multiple measures of model fit, configural, metric, and scalar invariance were found for almost all items across the age and gender groups, as was invariance in the factor covariances. The PedsQL^TM Multidimensional Fatigue Scale demonstrated strict factorial invariance for child and adolescent self-report across gender and strong factorial invariance across age subpopulations.

Conclusions The findings support an equivalent three-factor structure across the age and gender groups studied. Based on these data, it can be concluded that pediatric patients across the groups interpreted the items in a similar manner regardless of their age or gender, supporting the multidimensional factor structure interpretation of the PedsQL^TM Multidimensional Fatigue Scale.

The PedsQL^TM is available at http://www.pedsql.org.

J. W. Varni
Department of Pediatrics, College of Medicine, Texas A&M University, College Station, TX, USA

J. W. Varni (✉)
Department of Landscape Architecture and Urban Planning, College of Architecture, Texas A&M University, 3137 TAMU, College Station, TX 77843-3137, USA
e-mail: jvarni@arch.tamu.edu

A. A. Beaujean
Department of Educational Psychology, Baylor University, Waco, TX, USA

C. A. Limbers
Department of Psychology and Neuroscience, Baylor University, Waco, TX, USA

Introduction

Children are uniquely positioned to self-report their perspectives on their health and well-being through their perceptions of their health-related quality of life (HRQOL) outcomes. The last 15 years have evidenced a significant increase in the development and utilization of pediatric HRQOL and symptom-specific measures in an effort to improve pediatric patient health and determine the value of

health care services [1]. Although the measurement of pediatric self-reported HRQOL in clinical trials has been advocated for a number of years [2], the emerging paradigm shift toward patient-reported outcomes (PROs) has provided the opportunity to further emphasize the value for child self-report HRQOL measurement as efficacy outcomes in clinical trials [3, 4].

Age groups and health-related quality of life measurement

Recent US Food and Drug Administration guidelines recommend that instrument development and validation testing for children and adolescents be conducted within fairly narrow age groupings, and to determine the lower age limit at which children can provide reliable and valid responses that can be compared across age categories [4]. Consistent with these recommendations, it has been an explicit goal of the PedsQL^{TM} Measurement Model to develop and test brief measures for the broadest age group empirically feasible, specifically including pediatric patient self-report for the youngest children possible [1, 5]. The PedsQL^{TM} scales include child self-report for aged 5–18 and parent proxy-report for aged 2–18 [6]. The items chosen for inclusion were initially derived from the measurement properties of the child self-report scales, while the parent proxy-report scales were constructed to directly parallel the child self-report items. Thus, the development and testing of the PedsQL^{TM} as a pediatric PRO explicitly emphasizes the child's perceptions, including children as young as 5 years of age [7], and consequently serves as an age-appropriate instrument to test the lower age limits achievable for factorial invariance of child self-report.

Gender differences and health-related quality of life measurement

Gender differences in health outcomes have been extensively documented in children and adolescents [8]. In children, adolescents, and young adults, gender differences in self-reported HRQOL have been demonstrated irrespective of the instrument utilized [9–13]. However, in order to have greater confidence that a HRQOL instrument is measuring the same constructs across different gender groups (i.e., that the items have the same meaning for boys and girls), it is essential to demonstrate measurement invariance across gender [14, 15].

Factorial invariance

Generic HRQOL and symptom-specific instruments enable comparisons across diverse pediatric populations [5, 16]. In order for these comparisons to be valid, items on such an instrument must have equivalent meaning across the sub-populations being compared [17, 18], that is, they must demonstrate factorial invariance [19]. Multigroup confirmatory factor analysis (MG-CFA) is one method used to assess these levels of factorial invariance across groups [17, 19]. To the degree that the components of the factor model (i.e., factor patterns, intercepts, and covariances) are determined to be equal across subpopulations, factorial invariance of an instrument can be inferred [17].

While the use of MG-CFA for invariance testing has grown substantially in recent years, there has been a relative absence of studies that have examined the factorial invariance of symptom-specific measures particularly in pediatric populations. Further, most of these studies have focused on establishing *configural* and *metric* invariance, ignoring higher levels of invariance that assess group differences in item-specific intercepts (i.e., strong/scalar or strict invariance) [19]. Without establishing stricter levels of invariance, age and gender differences in scores on health outcome instruments may be confounded by differences in what the instrument is measuring among the groups [20]. Table 1 shows the levels of invariance, the constraints imposed by the level, and the between-group comparisons allowed if the invariance level is tenable.

Recently, studies have demonstrated the factorial invariance of the PedsQL^{TM} 4.0 Generic Core Scales longitudinally and across a number of subpopulations [21–26], including age and gender [27, 28]. However, the factorial invariance of the fatigue construct utilizing the PedsQL^{TM} Multidimensional Fatigue Scale across age and gender has not been previously reported.

The PedsQL^{TM} Multidimensional Fatigue Scale

The PedsQL^{TM} Multidimensional Fatigue Scale was designed as a generic symptom-specific instrument to measure fatigue across pediatric populations [29, 30]. Previously, the PedsQL^{TM} Multidimensional Fatigue Scale scores have demonstrated good to excellent pediatric patient self-report reliability and validity in pediatric patients across a number of pediatric chronic health conditions, demonstrating significant correlations with generic HRQOL (more fatigue symptoms associated with lower generic HRQOL) [29–39], including an expanding number of international translations and studies with young adult and adult patients [40–43]. However, we are not aware of a study which has used the Multigroup CFA framework to compare the factorial invariance of the PedsQL^{TM} Multidimensional Fatigue Scale, which may have significant utility as a statistical method for international cross-cultural assessment research in which different age and gender and other subpopulations are compared within and across countries.

Table 1 Levels of measurement invariance from least to most restrictive

Model	Title	Description	Comparisons allowed between groups
1	Configural	The factor model for all groups is the same. No parameter constraints are imposed	None
2	Weak/metric	1 + all factor pattern coefficients constrained to be equal between groups (but can vary within a group). Factor (co)variances allowed to vary between groups	Factor (co)variances [weak evidence]
3	Strong/scalar	2 + all intercepts are constrained to be equal between groups (but can vary within a group). Factor means allowed to vary between groups	Factor means, factor (co)variances [strong evidence]
4	Strict	3 + measurement error variances are constrained to be the same between groups (but can vary within a group)	
5	Homogeneity of construct variance	Constrain the factor variances to be the same across groups	Construct reliability

Consequently, the objective of the present study was to examine the factorial invariance of child self-reported multidimensional fatigue across three age groups and gender utilizing the Multigroup CFA framework with the PedsQL[TM] Multidimensional Fatigue Scale.

Methods

Participants and settings

The sample contains item-level data from previously published child self-reports ($n = 837$) [29, 30, 33, 35–37]. Participants were recruited from general pediatric clinics, subspecialty clinics, and hospitals in which children were being seen for well-child checks, mild acute illness, or chronic illness care. Participants were assessed in person or by telephone. For in-person mode of administration, research assistants obtained written parental informed consent and child assent. Paper-and-pencil surveys were self-administered for children aged 8–18 and interview-administered for children aged 5–7 (and also in situations where the child was unable to read or write as a consequence of either physical or cognitive impairment). For telephone administration, parents of children aged 5–18 were called by a research assistant who explained the study and obtained verbal parental informed consent and child assent. The research assistant verbally administered the PedsQL[TM] Multidimensional Fatigue Scale to the child. If the child was not home at the time of the initial call, the research assistant arranged for a call at another time. These research protocols were approved by all appropriate Institutional Review Boards (IRBs).

The average age of the 379 girls (45.3 %) and 377 boys (45.0 %) was 12.32 ± 4.98 years. Gender data were missing for 81 participants (9.7 %). With regard to race/ethnicity, the sample contained 308 (36.8 %) White

non-Hispanic, 195 (23.3 %) Hispanic, 58 (6.9 %) Black non-Hispanic, 25 (3.0 %) Asian/Pacific Islander, 71 (8.5 %) American Indian or Alaskan Native, and 40 (4.8 %) other. Race/ethnicity data were missing for 140 participants (16.7 %). Seventy-five percent of the sample had a chronic health condition.

Missing survey data

On the 18-item PedsQL[TM] Multidimensional Fatigue Scale, one or more item responses were missing for 38 (5 %) of the respondents. Of those 38, 28 (74 %) were missing only one response, 7 (18 %) were missing only two responses, 1 (3 %) was missing 4 responses, 1 (3 %) was missing 8 responses, and 1 (3 %) was missing all responses. The respondent with missing values for all items was removed from the analysis and the other missing data were handled by using full information maximum likelihood estimation [44].

Measures

The PedsQL[TM] Multidimensional Fatigue Scale

The PedsQL[TM] Multidimensional Fatigue Scale (MFS) is an 18-item instrument encompassing three scales: (1) General Fatigue (6 items, e.g., "I feel tired."; "I feel too tired to do things that I like to do."), (2) Sleep/Rest Fatigue (6 items, e.g., "I feel tired when I wake up in the morning."; "I rest a lot."), and (3) Cognitive Fatigue (6 items, e.g., "It is hard for me to keep my attention on things."; "It is hard for me to remember what people tell me"). The PedsQL[TM] MFS was developed based on the research and clinical experiences in pediatric chronic health conditions, and the instrument development literature [45–47], which consisted of a review of the extant literature on fatigue in both adult and pediatric patients, patient and parent focus

groups and individual focus interviews, item generation, cognitive interviewing, pretesting, and subsequent field testing of the new measurement instrument [29, 30]. In addition to the original studies [29, 30], subsequent studies have continued to support the reliability and validity of the PedsQL[TM] MFS scores for use with children and adolescents diagnosed with multiple chronic health conditions [29–36, 39, 42, 43, 48, 49].

The format, instructions, response scale, and scoring method are identical to the PedsQL[TM] 4.0 Generic Core Scales, with higher scores indicating lower fatigue symptoms. The instructions ask how much of a problem each item has been either during the past month or the past 7 days. The PedsQL[TM] MFS has two forms, a child self-report and a parent proxy-report. There are three separate child and adolescent self-report forms: ages 5–7, 8–12, and 13–18 years. Only child self-report was utilized in the present study. The self-report forms were designed to be parallel, with the items only differing in reading level. A 5-point Likert-type response scale is utilized for the 8–12 and 13–18-year-old forms (0 = never a problem; 1 = almost never a problem; 2 = sometimes a problem; 3 = often a problem; 4 = almost always a problem). The 5–7-year-old form was simplified to a 3-point scale to increase the ease of use (0 = not at all a problem; 2 = sometimes a problem; 4 = a lot of a problem), with each response choice anchored to a happy to sad faces scale. Items are reverse-scored and linearly transformed to a 0–100 scale (0 = 100, 1 = 75, 2 = 50, 3 = 25, 4 = 0), so that higher scores indicate better HRQOL (i.e., fewer symptoms of fatigue). Scale scores are computed as the sum of item scores divided by the number of answered items. If more than 50 % of the items in the scale are missing, the Scale score is not computed [50].

PedsQL[TM] Family Information Form

The PedsQL[TM] Family Information Form [6] is a demographic questionnaire for parents to complete that asks about the child's date of birth, gender, and race/ethnicity.

Statistical analysis

Multiple group factor analysis

The purpose of this study was to examine the invariance of the PedsQL[TM] MFS items for the self-report forms across both age and gender. To assess invariance, we used a MG-CFA approach, which assesses the invariance of measurement parameters (e.g., factor patterns) across two or more groups by using a series of increasingly stringent, nested models (see Table 1) [51]. We tested model fit with a bifactor model of the items [52–54], which posits one general factor influencing all the items and three domain-specific factors representing the General, Sleep/Rest, and Cognitive Fatigue domains.

Researchers [20, 55] suggest two sets of criteria when testing for factorial invariance. The first ("traditional perspective") examines the change in chi-square values ($\Delta \chi^2$) across nested models. If, as the models grow more restrictive, the $\Delta \chi^2$ values do not "significantly" change (using a given α level), this is evidence that a more restrictive model fits the data as well as the less restrictive model; thus, the more restrictive (i.e., more parsimonious) model should be favored over the less restrictive one.

The use of $\Delta \chi^2$ values has been criticized because of the sensitivity to sample size [56]. Recently, Cheung and Rensvold [56] and Meade et al. [57] have argued that some alternative fit indices were not prone to this problem. Specifically, they found that the Comparative Fit Index (CFI) [58] and McDonald's [59] Noncentrality Index (NCI) were more robust across a variety of sample sizes. Thus, the second line of evaluations criteria ("practical perspective") recommends that invariance can be based on two criteria: (a) the multigroup factor model exhibits an adequate fit to the data and (b) the change in values for fit indices (e.g., Δ CFI, Δ NCI) is negligible.

Based on Byrne and Stewart's [55] and Little's [20] recommendation, this study used two sets of fit indices: one to assess overall model fit and the other to assess change in model fit between two models. As Hu and Bentler [60] recommend, we used multiple fit indices for both. For overall model fit, we included the root mean square error of approximation (RMSEA) [61], Comparative Fit Index (CFI) [62], McDonald's Noncentrality Index (NCI) [63], and the standardized root mean square residual (SRMR) [64]. These indices were chosen as they represent a variety of fit criteria and they tend to perform well in evaluating different models [65]. For both overall model fit as well as change in model fit, we looked for patterns in the fit statistics and judged acceptance/rejection of the specific model based on the majority of the indices. For this study's criteria of overall model data fit, we used the following: (a) RMSEA $\leq$0.08 [61, 66]; (b) SRMR $\leq$0.08, [60, 67]; (c) CFI $\geq$0.96 [68]; and (d) NCI $\geq$0.90 [67, 69].

To test the change in fit between nested models, we used the Δ CFI and Δ NCI. Cheung and Rensvold [56] suggested 0.01 as the threshold of Δ CFI and 0.02 as the threshold for Δ NCI. Meade et al. [57], however, suggested more restrictive values of 0.002 for Δ CFI and 0.007 for Δ NCI (based on having 3 factors and 18 indicators, p. 586) to maximize power. As this issue is not yet resolved, we considered both values for the Δ CFI and Δ NCI, with values less than Meade et al.'s [57] criteria indicating stronger evidence of invariance than values only meeting Cheung and Rensvold's [56] criteria. All analysis was done

in R [70] using the lavaan [71] and psych [72] statistical packages.

Analyzing item-level data

Before testing for invariance, the data were inspected to see whether a categorical data model or a continuous data model should be used. Previous research has shown that treating categorical variables as continuous is usually not problematic when the variables have at least three categories and do not have substantial differential skew (i.e., one variable is highly skewed positively and another is highly skewed negatively) [73–75].

To assess for skew, we first examined all items together using Mardia's test [76]. The results (b1p = 48.46, $\chi^2 = 6,453.78$, $df = 1,140$) indicate that some of the items might have substantial skew, so we examined each item independently. The only items with a skew statistics substantially above 1 were the fourth item on the General Fatigue Scale (skew = 1.74) and the fifth item on the Sleep/Rest Fatigue Scale (skew = 1.52). We subsequently plotted the frequency of each items response to examine whether the skew was in the same direction or different directions. Figures 1, 2, 3 indicate that any skew that exists is all in the same direction.

Next, we compared the correlation matrices between categorical (polychoric) and continuous (Pearson) estimators. The difference was minimal for the items within a domain (SRMR = 0.06) as well as items between domains (SRMR = 0.05). Second, separate exploratory factor analysis was conducted assuming the indicators were either continuous or categorical. Both the parallel analysis [77] and the minimum average partial analysis [78] (using both Pearson and the polychoric correlations) [79, 80] indicated that three factors should be extracted. Consequently, the indicators were treated as continuous for the purpose of this study, but we used a maximum likelihood estimator with robust standard errors [81], which is a better estimator to use when the data may not meet the multivariate normal assumption of traditional ML estimation [82], and help the performance of full information estimation [83].

Results

General model

Initially, we fit a baseline model using all the groups combined for each domain separately. Within each domain, there was a single factor that accounted for all the item covariance. For the General (Model General 1 in Table 2), Sleep/Rest (Model Sleep/Rest 1), and Cognitive (Model Cognitive 1) domains, the model did not fit the data badly, indicating that, within a domain, the items appear to be unidimensional. We then fit a model combing items from all three domains. First, we fit a combined model (Model All 1) that had the six items within a domain as the sole indicators of their intended factor and allowed the three domain factors to covary (i.e., oblique factors). According to most of the alternative fit indices, the model fits the data adequately, but the NCI was below the suggested value of 0.90. We then fit a bifactor model of the items [52, 53], which posits one general factor influencing all the items and three domain-specific factors representing the General, Sleep/Rest, and Cognitive Fatigue domains (Model All 2). All the alternative fit indices indicate that this model fits the data better than the three oblique factor model. Consequently, we then used the bifactor model (Model O2) as the baseline model to test for invariance. For a graphical representation of the model, see Fig. 4.

Gender

To examine invariance across gender, we split the data by gender. Eighty-one of the respondents' parents did not indicate the child's gender, so were not used for this analysis, leaving 756 in the data set. First, we assessed for configural invariance (see Model S1 in Table 3). The results suggest that the model fits the data relatively well, although the CFI and NCI values were on the border of the "acceptable" range.

The next step involved assessing for metric invariance, which we did by examining whether the factor pattern coefficients were the same across both genders. The results (Model S2) indicate that the overall model fits the data slightly better than model S1. Thus, there was enough evidence to continue the invariance assessment.

In the next model (Model S3), we examined scalar invariance by constraining all the indicator variables' intercepts (i.e., the scales' origins) across groups. The results suggest that the model fits the data relatively well, and the Δ CFI and Δ NCI values met both the Cheung and Rensvold's [56] and Meade et al.'s [57] criteria. Consequently, we tested for further levels of invariance.

The next step involved examining whether any of the subtests' unique (residual) variances (Model S4) were invariant across gender. While such invariance is not required to compare the latent constructs between male and females, it is a necessary step (along with the invariance of the factor variances) to determine whether the constructs' reliabilities are the same across groups [19]. The results indicated that this model fits the data fairly well, although the CFI and NCI values drop below the "acceptable" range. The Δ CFI and Δ NCI meet the Cheung and Rensvold [56] criteria, but not the Meade et al. [57] criteria. Consequently, we tested homogeneity of factor variances using both the scalar (Model S3) and strict (Model S4) models as the baseline (Table 4).

Fig. 1 Item response distributions for General Fatigue Scale
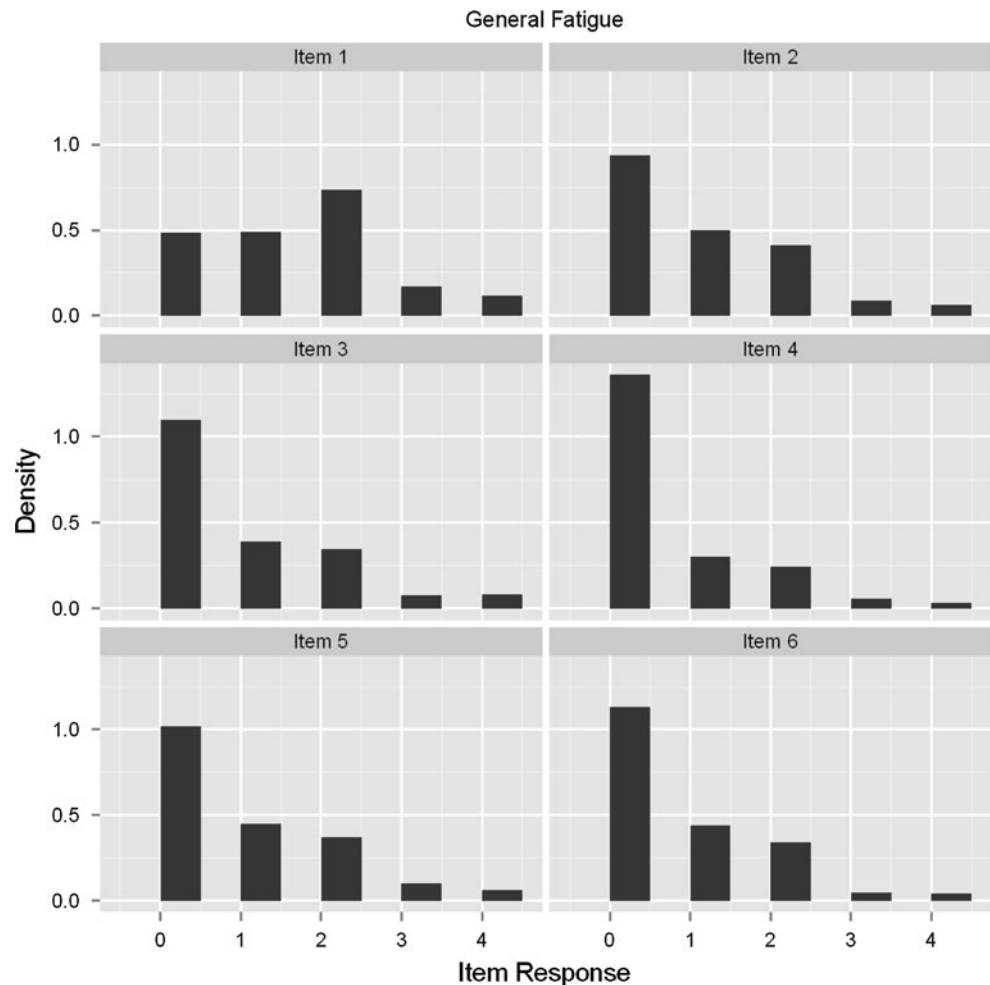
For the last step, we assessed whether the factor variances were invariant across the two groups (Models S5a and S5b). As both models fit the data as well as Model S3 and Model S4, respectively, the results indicate that the latent variances are not substantially different between genders. Subsequently, we assessed the reliability of each latent construct using ω [84]. Using the scalar invariance model, which allows different residual variances for males and females, the Sleep/Rest and Cognitive Fatigue Scales' construct reliabilities are 0.99 for both males and females. For the General Fatigue Scale, the construct reliability is 0.98 for males and 0.97 for females. For the general fatigue factor (i.e., the factor related to all fatigue items), the construct reliability was 0.99 for males and females. Using the strict invariance model, the construct reliabilities were 0.99 for all three of the domain factors, as well as for the general fatigue factor.

Age

To assess for invariance across age, we then split the data by age form: Young Child (5–7 years), Child (8–12 years),

and Adolescent (13–18 years). There were 87 participants (10.4 %) who completed the Young Child (5–7) form, 343 participants (41.0 %) who completed the Child (8–12) form, and 407 participants (48.6 %) who completed the Adolescent (13–18) form.

First, we assessed for configural invariance (see Model A1 in Table 5). All the alternative fit indices except NCI indicate that the model does not fit the data badly. Consequently, we used it to test the subsequent invariance model.

The next step involved assessing for metric invariance, which we did by examining whether the factor pattern coefficients were the same across age groups. The results (Model A2) indicate that the overall model fits the data no worse than model A1, and the Δ CFI and Δ NCI values met the Cheung and Rensvold [56] criteria. Thus, there appeared to be enough evidence to continue the invariance assessment.

In the next model (A3), we examined scalar invariance by constraining all the indicator variables' intercepts (i.e., the scales' origins) across age groups. The evidence is mixed on how the model fits the data. The RMSEA and SRMR indicate that the model does not fit the data badly,

Fig. 2 Item response distributions for Sleep/Rest Scale
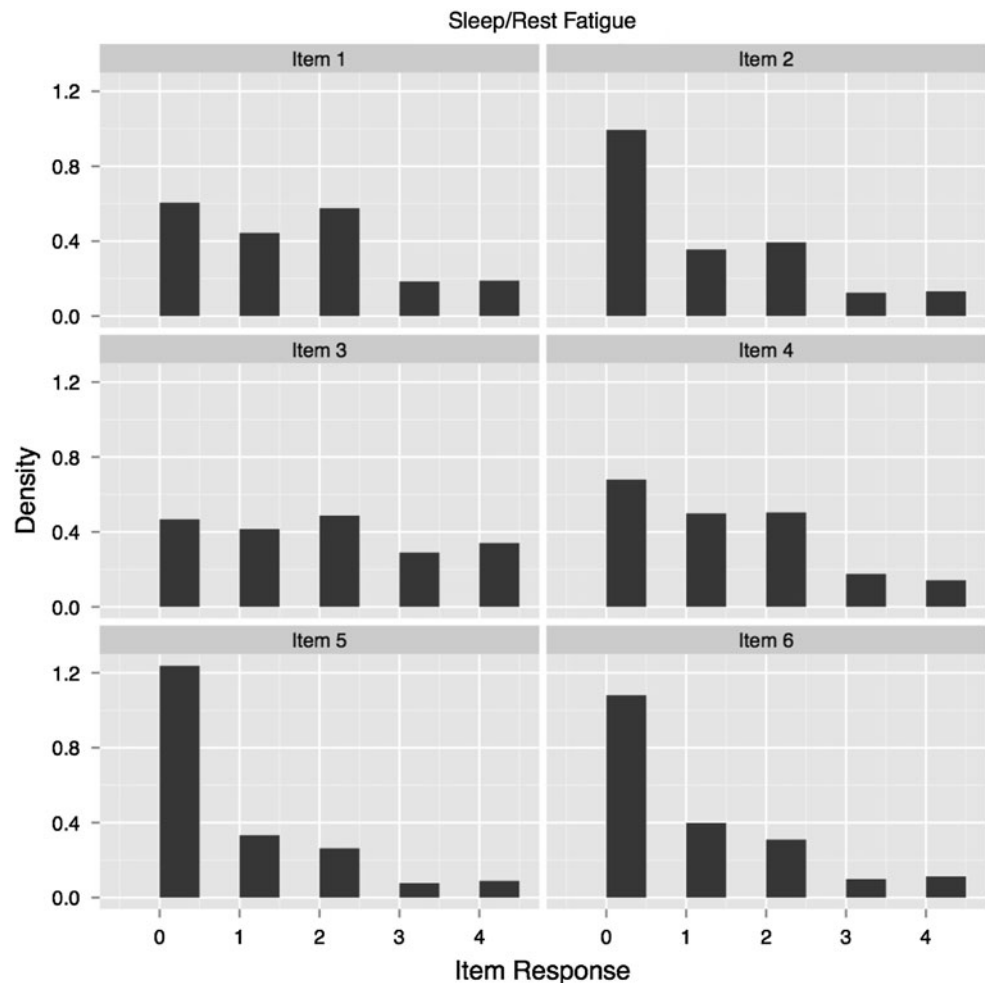
but the NCI did not indicate a good fit to the data and the CFI is on the border of the unacceptable range. The Δ CFI met the Cheung and Rensvold [56] criteria, but the Δ NCI values did not. Consequently, we examined the intercepts in model A2 to determine whether there were any large differences in intercepts among the groups. The differences were minimal for a single intercept, so the model A3 was kept for future model assessment (Table 6).

The next step involved examining whether any of the subtests' unique (residual) variances (Model A4) were invariant across the age groups. Such invariance is not required to compare the constructs across aged, but it is a necessary step (along with invariance of the factor variances) to determine whether the constructs' reliabilities are the same across groups [19]. All fit indices for this model indicate that it did not fit the data well. Further data inspection showed that the majority of the residual variances needed to be unconstrained for this model to fit well, so none of them were considered to be equivalent across the age forms.

We next examined whether the factor variances were invariant across the age groups (Model A5). The results indicate this model fits the data as well as Model A3, with

the Δ CFI meeting the Cheung and Rensvold [56] criteria and the Δ NCI meeting both the Cheung and Rensvold [56] and Meade et al. [57] criteria, so the variances were left equal across the age forms.

Subsequently, we assessed the reliability of each latent construct using ω [84] using model A5, which allows different residual variances for the age groups, but constrains the latent variables' variances to be the same across groups. For the Sleep/Rest and Cognitive Fatigue Scales, construct reliability is 0.98 for the Young Child form and 0.99 for the Child and Adolescent forms. For the General Fatigue Scale, the construct reliability is 0.97 for the Young Child form and 0.99 for the Child and Adolescent forms. For the general fatigue factor (i.e., the factor related to all fatigue items), the construct reliability was 0.98 for the Young Child form and 0.99 for the for the Child and Adolescent forms.

Discussion

The present findings demonstrate that when self-reporting their fatigue, pediatric patients who completed the PedsQL[TM]

Fig. 3 Item response
distributions for Cognitive Scale
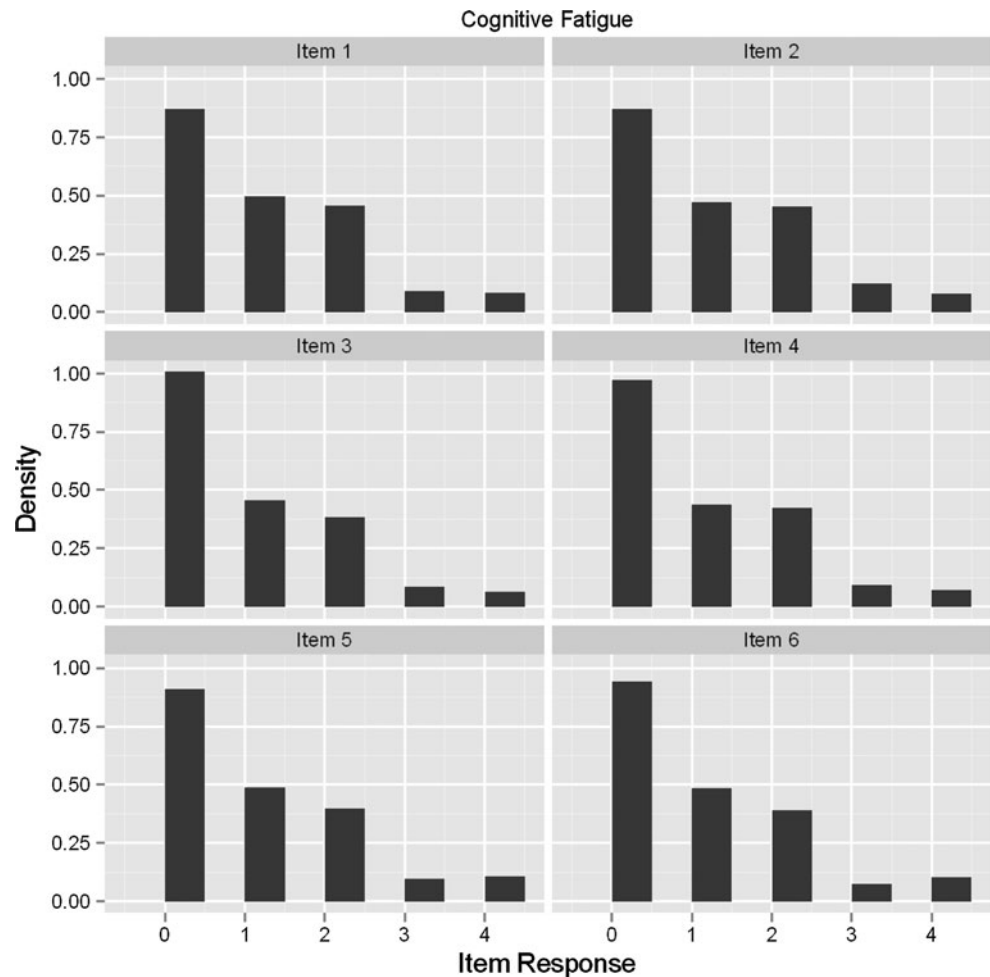
Table 2 Results from confirmatory factor analysis of PedsQL[TM] Multidimensional Fatigue Scale using all respondents

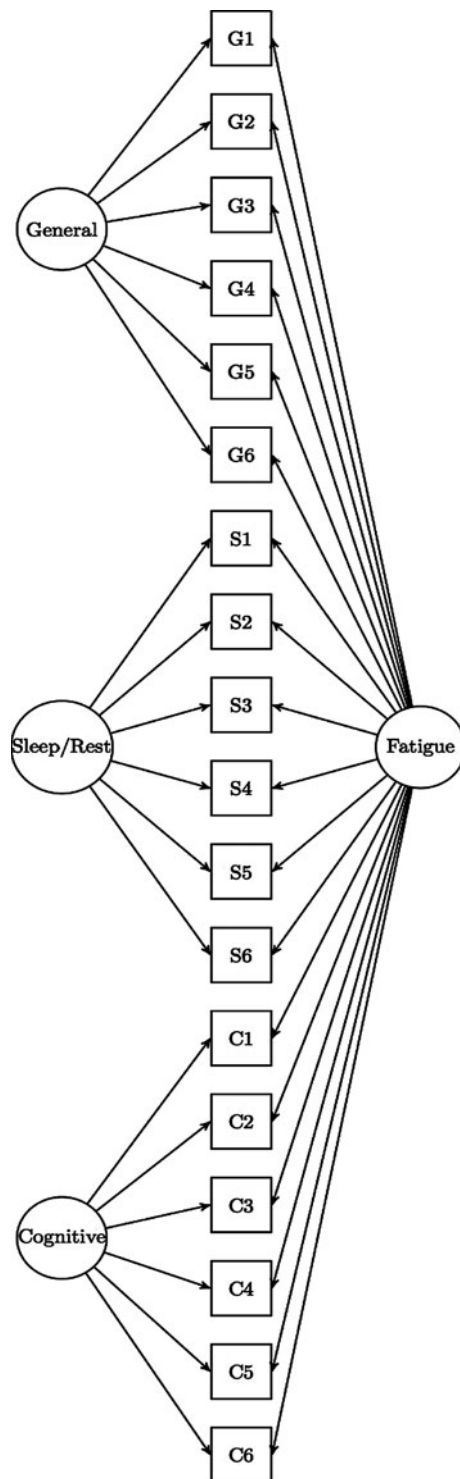| Model	Model description	χ^{2*}	df	p	Scaling	RMSEA	CFI	SRMR	NCI
General 1	General fatigue	59.74	9	0.00	1.97	0.08	0.95	0.04	0.97
Sleep/Rest 1	Sleep/rest fatigue	61.33	9	0.00	1.38	0.08	0.95	0.04	0.97
Cognitive 1	Cognitive fatigue	17.41	9	0.04	2.00	0.03	0.99	0.02	0.99
All 1	3-Factor, oblique	438.01	132	0.00	1.57	0.05	0.93	0.06	0.83
All 2	3-Factor, bifactor	250.23	117	0.00	1.55	0.04	0.97	0.03	0.92

χ^{2*}: Scaled χ^2, *RMSEA* root mean square error of approximation, *CFI* Comparative Fit Index, *SRMR* standardized root mean square residual, *NCI* McDonald's Noncentrality Index

Multidimensional Fatigue Scale (MFS) across the three age groups studied and gender had a similar three-factor multi-dimensional fatigue model structure, that is, across gender and age groupings, the MFS items related to their intended constructs similarly and had similar conditional means (intercepts). The factor covariances were the same across gender and age groupings, as were the factor mean scores, except for the cognitive domain across the age groups. Thus, across both gender and age, the MFS scores have the same meaning and can be interpreted similarly. While the residual variances were similar between males and

females, they were not among the age groupings. Nonetheless, the reliability estimates were high across all groupings ($\omega \geq 0.97$), meaning that while the reliability of scores might differ among the age forms, the difference is minimal and, in general, the variability among observed scores is largely due to variance among the constructs they are designed to measure.

Raju et al. [85] succinctly describe the importance of measurement equivalence by stating that "When measurement equivalence is present, the relationship between the latent variable and the observed variable remains

Fig. 4 Bifactor model of PedsQL[TM] Multidimensional Fatigue Scale

latent constructs across groups, including gender and age subpopulations. MG-CFA across age and gender subpopulations has previously not been conducted in fatigue measurement in pediatric patients utilizing child and adolescent self-reported multidimensional fatigue instruments.

The MG-CFA statistical methods utilized in the present study have important implications in general for international comparative clinical research in children and adolescents in which different age and gender subpopulations are studies. Standardized assessment instruments must demonstrate that test items are interpreted similarly across age, gender, language, socioeconomic, health status, and race/ethnicity subpopulations [86]. Demonstrating stricter levels of the hierarchy of factorial invariance across these subpopulations is critical given the growing importance of patient-reported outcomes in international clinical trials, health disparities analyses, and comparative health research. Since fatigue has been found to be a common symptom in adolescent populations [87, 88], with potential gender differences in fatigue reporting [87, 89], then the demonstration of factorial invariance across age and gender is an essential step in further understanding any differences in pediatric self-reporting of multidimensional fatigue symptoms, including fatigue associations with, for example, physical, emotional, social, and school functioning in pediatric populations. Imperative to the evaluation of interventions designed to reduce fatigue is the need for reliable and valid assessment of multidimensional fatigue in pediatric populations, particularly given that previous longitudinal research in general pediatric populations of adolescents indicates that persistently fatigued participants demonstrate higher level of depression and anxiety, are less physically active, and sleep a shorter duration at night [88, 89].

The present findings contribute to the empirical literature on the PedsQL[TM] Measurement Model by demonstrating strict factorial invariance for child and adolescent multidimensional fatigue self-report across gender and strong factorial invariance across age subpopulations. The results of the present study suggest that when mean differences are found utilizing the PedsQL[TM] Multidimensional Fatigue Scale across the three age and gender subpopulations studied, these differences are more likely real differences in self-perceived multidimensional fatigue, rather than differences in the interpretation of the items as a function of age and gender. To our knowledge, the present study represents the first empirical test of the multidimensional fatigue construct in pediatric populations utilizing a bifactor model while testing for factorial invariance. These findings have important implications for scoring the PedsQL[TM] Multidimensional Fatigue Scale and suggest that both a total scale score comprised of all 18 items, as well as individual scale scores for each of the

invariant across populations. In this case, the observed mean difference may be viewed as reflecting only the true difference between the populations" (p. 517). Factorial invariance is an essential component of the iterative process of demonstrating the measurement equivalence of

Table 3 Test for invariance by gender on PedsQL[TM] Multidimensional Fatigue Scale

Model	Description	χ^2	df	$\Delta \chi^2$	$p (\Delta \chi^2)$	RMSEA	CFI	Δ CFI	SRMR	NCI	Δ NCI
S1	Same structure, but all parameters freely estimated[a]	397.91	234	–	–	0.04	0.956	–	0.04	0.897	–
S2	S1 + loadings equal (free variances, though)	412.89	266	36.79	0.26	0.04	0.961	*	0.05	0.907	*
S3	S2 + intercepts equal (free means, though)	430.60	280	22.22	0.07	0.04	0.960	0.001	0.05	0.905	0.002
S4	S3 + constrain unique variances	465.17	298	51.88	0.00	0.04	0.955	0.004	0.05	0.895	0.010
S5a	S3 + equal factor variance	434.62	284	6.10	0.19	0.04	0.960	0.000	0.05	0.905	0.000
S5b	S4 + equal factor variance	470.70	302	8.82	0.07	0.04	0.955	0.000	0.05	0.894	0.000

The p values for the χ^2 as well as the MLR scaling factors are not shown due to space considerations

All p values were <0.001 and all scaling factors were between 1.43 and 1.46. χ^2 and $\Delta \chi^2$ values represent the scaled values [91]

RMSEA root mean square error of approximation, *CFI*: Comparative Fit Index, *SRMR* standardized root mean square residual, *NCI* McDonald's Noncentrality Index

[a] Model *All 2* from Table 2 was used as the baseline model, only fit separately for the two gender groups

Table 4 Standardized estimates of intercepts and factor loadings for MG-CFA model by gender

Domain	Item	Domain factor loadings	SE	General factor loadings	SE	Intercept	SE
General	Item 1	0.29	0.08	0.76	0.05	1.62	0.06
General	Item 2	0.23	0.08	0.73	0.04	1.04	0.05
General	Item 3	0.18	0.10	0.81	0.04	0.95	0.05
General	Item 4	0.08	0.09	0.63	0.04	0.63	0.05
General	Item 5	−0.46	0.11	0.78	0.05	0.85	0.05
General	Item 6	−0.17	0.06	0.62	0.04	0.73	0.05
Sleep/Rest	Item 1	0.66	0.06	0.55	0.05	1.50	0.06
Sleep/Rest	Item 2	0.07	0.07	0.65	0.05	1.09	0.06
Sleep/Rest	Item 3	0.11	0.06	0.76	0.04	1.91	0.06
Sleep/Rest	Item 4	0.72	0.05	0.60	0.05	1.34	0.06
Sleep/Rest	Item 5	0.67	0.06	0.49	0.05	0.73	0.05
Sleep/Rest	Item 6	0.52	0.06	0.69	0.04	0.94	0.06
Cognitive	Item 1	0.33	0.05	0.66	0.05	1.07	0.05
Cognitive	Item 2	0.58	0.05	0.63	0.04	1.06	0.05
Cognitive	Item 3	0.58	0.05	0.56	0.05	0.88	0.05
Cognitive	Item 4	0.53	0.05	0.54	0.05	0.96	0.05
Cognitive	Item 5	0.56	0.05	0.55	0.05	1.04	0.05
Cognitive	Item 6	0.66	0.05	0.59	0.05	0.98	0.05

Parameter estimates are shown for Model S5a (see Table 3)

Table 5 Test for invariance by age form on PedsQL[TM] Multidimensional Fatigue Scale

Model	Description	χ^2	df	$\Delta \chi^2$	p	RMSEA	CFI	Δ CFI	SRMR	NCI	Δ NCI
A1	Same structure, but all parameters freely estimated[a]	535.49	357	–	–	0.04	0.965	–	0.04	0.899	–
A2	A1 + loadings equal (free variances, though)	606.32	415	87.81	0.01	0.04	0.963	0.002	0.06	0.892	0.007
A3	A2 + intercepts equal (free means, though)	677.47	443	83.72	0.00	0.04	0.955	0.008	0.06	0.869	0.023
A4	A3 + constrain unique variances	1,257.39	479	706.71	0.00	0.08	0.849	0.106	0.07	0.627	0.242
A5	A3 + Factor variances	693.79	451	14.27	0.08	0.04	0.953	0.002	0.07	0.865	0.004

The p values for the χ^2 as well as the MLR scaling factors are not shown because of space considerations. All p values were <0.001 and all scaling factors were between 1.26 and 1.30

$\Delta \chi^2$ values represent the scaled values [91]

RMSEA root mean square error of approximation, *CFI* Comparative Fit Index, *SRMR* standardized root mean square residual, *NCI* McDonald's Noncentrality Index

[a] Model *All 2* from Table 2 was used as the baseline model, only fit separately for the three age groups

Table 6 Standardized estimates of intercepts and factor loadings for MG-CFA model by age

Domain	Item	Domain factor loadings	SE	General factor loadings	SE	Intercept	SE
General	Item 1	0.25	0.09	0.73	0.04	1.40	0.05
General	Item 2	0.24	0.07	0.76	0.04	0.84	0.05
General	Item 3	0.53	0.20	0.76	0.09	0.70	0.05
General	Item 4	0.37	0.16	0.59	0.08	0.46	0.04
General	Item 5	−0.08	0.29	0.76	0.08	0.78	0.06
General	Item 6	0.00	0.23	0.63	0.07	0.65	0.05
Sleep/Rest	Item 1	0.61	0.07	0.55	0.06	1.35	0.06
Sleep/Rest	Item 2	0.03	0.09	0.67	0.06	0.95	0.05
Sleep/Rest	Item 3	0.12	0.11	0.79	0.09	1.72	0.06
Sleep/Rest	Item 4	0.71	0.05	0.61	0.06	1.16	0.06
Sleep/Rest	Item 5	0.65	0.06	0.48	0.05	0.60	0.05
Sleep/Rest	Item 6	0.51	0.06	0.69	0.05	0.75	0.05
Cognitive	Item 1	0.26	0.06	0.68	0.05	0.97	0.05
Cognitive	Item 2	0.55	0.07	0.65	0.05	1.07	0.05
Cognitive	Item 3	0.53	0.05	0.59	0.05	0.89	0.05
Cognitive	Item 4	0.55	0.04	0.55	0.04	0.93	0.05
Cognitive	Item 5	0.53	0.05	0.58	0.05	0.99	0.05
Cognitive	Item 6	0.60	0.04	0.59	0.04	0.96	0.05

Parameter estimates are shown for Model A5 (see Table 5)

6-item individual scales (General Fatigue Scale, Sleep/Rest Fatigue Scale, and Cognitive Fatigue Scale) are justified [90].

The present study has several limitations. First, given the available sample size, the age groups were limited to the PedsQL™ instrument age groups of 5–7, 8–12, and 13–18 years. With a larger sample size, it would have been more ideal to study each individual age group, as well as race/ethnicity and mode of administration subgroups, as we were able to do for the PedsQL™ 4.0 Generic Core Scales factorial invariance analyses [23, 24, 27]. Further, the sample size for the 5–7 age group would have ideally been larger than what was available in the existing database in order to increase statistical power. Working from an existing database, information on nonparticipants was not available nor were response rates from the groups studied. Also, there were missing age and gender information for some of the participants which may limit the generalizability of the findings. Nevertheless, these findings complement the previous findings that the PedsQL™ 4.0 Generic Core Scales demonstrate factorial invariance for child and adolescents self-report across age, gender, language, socioeconomic, health status, and race/ethnicity subpopulations [21–23, 26–28], and add factorial invariance analyses across age and gender for the PedsQL™ Multidimensional Fatigue Scale to this emerging list of pediatric factorial invariance studies in HRQOL and symptom-specific instruments.

References

1. Varni, J. W., & Limbers, C. A. (2009). The Pediatric Quality of Life Inventory™: Measuring pediatric health-related quality of life from the perspective of children and their parents. *Pediatric Clinics of North America, 56*, 843–863.
2. Eiser, C. (2004). Use of quality of life measures in clinical trials. *Ambulatory Pediatrics, 4*, 395–399.
3. Varni, J. W., Handen, B. L., Corey-Lisle, P. K., Guo, Z., Manos, G., Ammerman, D. K., et al. (2012). Effect of Aripiprazole 2 to 15 mg/d on health-related quality of life in the treatment of irritability associated with Autistic Disorder in children: A post hoc analysis of two controlled trials. *Clinical Therapeutics, 34*, 980–992.
4. FDA. (2009). Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Food and Drug Administration, U.S. Department of Health and Human Services, Rockville, MD.
5. Varni, J. W., Seid, M., & Rode, C. A. (1999). The PedsQL™: Measurement model for the Pediatric Quality of Life Inventory. *Medical Care, 37*, 126–139.
6. Varni, J. W., Seid, M., & Kurtin, P. S. (2001). PedsQL™ 4.0: Reliability and validity of the Pediatric Quality of Life Inventory™ version 4.0 Generic Core Scales in healthy and patient populations. *Medical Care, 39*, 800–812.
7. Varni, J. W., Limbers, C. A., & Burwinkle, T. M. (2007). How young can children reliably and validly self-report their

health-related quality of life?: An analysis of 8,591 children across age subgroups with the PedsQL™ 4.0 Generic Core Scales. *Health and Quality of Life Outcomes, 5*(1), 1–13.

8. Sweeting, H. (1995). Reversals of fortune?: Sex differences in health in childhood and adolescence. *Social Science and Medicine, 40*, 77–90.

9. Arrington-Sanders, R., Yi, M. S., Tsevat, J., Wilmott, R. W., Mrus, J. M., & Britto, M. T. (2006). Gender differences in health-related quality of life of adolescents with cystic fibrosis. *Health and Quality of Life Outcomes, 4*(5), 1–8.

10. Jorngarden, A., Wettergen, L., & von Essen, L. (2006). Measuring health-related quality of life in adolescents and young adults: Swedish normative data for the SF-36 and the HADS, and the influence of age, gender, and method of administration. *Health and Quality of Life Outcomes, 4*(91), 1–10.

11. Reinfjell, T., Diseth, T. H., Veenstra, M., & Vikan, A. (2006). Measuring health-related quality of life in young adolescents: Reliability and validity in the Norwegian version of the Pediatric Quality of Life Inventory™ 4.0 (PedsQL™) Generic Core Scales. *Health and Quality of Life Outcomes, 4*(61), 1–9.

12. Upton, P., Eiser, C., Cheung, I., Hutchings, H. A., Jenney, M., Maddocks, A., et al. (2005). Measurement properties of the UK-English version of the Pediatric Quality of Life Inventory™ 4.0 (PedsQL™) Generic Core Scales. *Health and Quality of Life Outcomes, 3*(22), 1–7.

13. Vaez, M., & Laflamme, L. (2003). Health behaviors, self-rated health, and quality of life: A study among first-year Swedish university students. *Journal of American College Health, 51*, 156–162.

14. Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*, S69–S77.

15. Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups?: Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44*, S78–S94.

16. Varni, J. W., Stucky, B. D., Thissen, D., DeWitt, E. M., Irwin, D. E., Lai, J. S., et al. (2010). PROMIS Pediatric Pain Interference Scale: An item response theory analysis of the pediatric pain item bank. *Journal of Pain, 11*, 1109–1119.

17. French, B., & Finch, W. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling, 13*, 378–402.

18. Teresi, J. A. (2006). Overview of quantitative measurement methods: Equivalence, invariance, and differential item functioning in health applications. *Medical Care, 44*, S39–S49.

19. Vandenberg, R., & Lance, C. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.

20. Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53–76.

21. Limbers, C. A., Newman, D. A., & Varni, J. W. (2008). Factorial invariance of child self-report across healthy and chronic health condition groups: A confirmatory factor analysis utilizing the PedsQL™ 4.0 Generic Core Scales. *Journal of Pediatric Psychology, 33*, 630–639.

22. Limbers, C. A., Newman, D. A., & Varni, J. W. (2008). Factorial invariance of child self-report across socioeconomic status groups: A multigroup confirmatory factor analysis utilizing the PedsQL™ 4.0 Generic Core Scales. *Journal of Behavioral Medicine, 31*, 401–411.

23. Limbers, C. A., Newman, D. A., & Varni, J. W. (2009). Factorial invariance of child self-report across race/ethnicity groups: A multigroup confirmatory factor analysis approach utilizing the PedsQL™ 4.0 Generic Core Scales. *Annals of Epidemiology, 19*, 575–581.

24. Varni, J. W., Limbers, C. A., & Newman, D. A. (2009). Using factor analysis to confirm the validity of children's self-reported health-related quality of life across different modes of administration. *Clinical Trials, 6*, 185–195.

25. Varni, J. W., Limbers, C. A., Newman, D. A., & Seid, M. (2008). Longitudinal factorial invariance of the PedsQL™ 4.0 Generic Core Scales child self-report version: One year prospective evidence from the California State Children's Health Insurance Program (SCHIP). *Quality of Life Research, 17*, 1153–1162.

26. Newman, D. A., Limbers, C. A., & Varni, J. W. (2010). Factorial invariance of child self-report across English and Spanish language groups in a Hispanic population utilizing the PedsQL™ 4.0 Generic Core Scales. *European Journal of Psychological Assessment, 26*, 194–202.

27. Limbers, C. A., Newman, D. A., & Varni, J. W. (2008). Factorial invariance of child self-report across age subgroups: A confirmatory factor analysis of ages 5 to 16 years utilizing the PedsQL™ 4.0 Generic Core Scales. *Value in Health, 11*, 659–668.

28. Varni, J. W., Limbers, C. A., & Newman, D. A. (2008). Factorial invariance of the PedsQL™ 4.0 Generic Core Scales child self-report across gender: A multigroup confirmatory factor analysis with 11,356 children ages 5 to 18. *Applied Research in Quality of Life, 3*, 137–148.

29. Varni, J. W., Burwinkle, T. M., Katz, E. R., Meeske, K., & Dickinson, P. (2002). The PedsQL™ in pediatric cancer: Reliability and validity of the Pediatric Quality of Life Inventory™ Generic Core Scales, Multidimensional Fatigue Scale, and cancer module. *Cancer, 94*, 2090–2106.

30. Varni, J. W., Burwinkle, T. M., & Szer, I. S. (2004). The PedsQL™ Multidimensional Fatigue Scale in pediatric rheumatology: Reliability and validity. *Journal of Rheumatology, 31*, 2494–2500.

31. Varni, J. W., Burwinkle, T. M., Limbers, C. A., & Szer, I. S. (2007). The PedsQL™ as a patient-reported outcome in children and adolescents with fibromyalgia: An analysis of OMERACT domains. *Health and Quality of Life Outcomes, 5*(9), 1–12.

32. Palmer, S. N., Meeske, K. A., Katz, E. R., Burwinkle, T. M., & Varni, J. W. (2007). The PedsQL™ Brain Tumor Module: Initial reliability and validity. *Pediatric Blood & Cancer, 49*, 287–293.

33. Marcus, S. B., Strople, J. A., Neighbors, K., Weissberg-Benchell, J., Nelson, S. P., Limbers, C. A., et al. (2009). Fatigue and health-related quality of life in pediatric inflammatory bowel disease. *Clinical Gastroenterology and Hepatology, 7*, 554–561.

34. Gold, J. I., Mahrer, N. E., Yee, J., & Palermo, T. M. (2009). Pain, fatigue, and health-related quality of life in children and adolescents with chronic pain. *Clinical Journal of Pain, 25*, 407–412.

35. Varni, J. W., Limbers, C. A., Bryant, W. P., & Wilson, D. P. (2009). The PedsQL™ Multidimensional Fatigue Scale in type 1 diabetes: Feasibility, reliability, and validity. *Pediatric Diabetes, 10*, 321–328.

36. Varni, J. W., Limbers, C. A., Bryant, W. P., & Wilson, D. P. (2010). The PedsQL™ Multidimensional Fatigue Scale in pediatric obesity: Feasibility, reliability, and validity. *International Journal of Pediatric Obesity, 5*, 34–42.

37. Varni, J. W., Limbers, C. A., Bryant, W. P., & Wilson, D. P. (2012). Assessment of fatigue in pediatric patients with short stature utilizing the PedsQL™ Multidimensional Fatigue Scale. *Children's Health Care, 41*, 162–181.

38. Pohl, J. F., Limbers, C. A., Kay, M. T., Harman, A., Rollins, M., & Varni, J. W. (2012). Health-related quality of life in pediatric patients with long-standing pancreatitis. *Journal of Pediatric Gastroenterology and Nutrition, 54*, 657–663.

39. MacAllister, W. S., Christodoulou, C., Troxell, R., Milazzo, M., Block, P., Preston, T. E., et al. (2009). Fatigue and quality of life in pediatric multiple sclerosis. *Multiple Sclerosis, 15*, 1502–1508.

40. Varni, J. W., & Limbers, C. A. (2008). The PedsQL™ Multidimensional Fatigue Scale in young adults: Feasibility, reliability and validity in a university student population. *Quality of Life Research, 17*, 105–114.

41. Robert, R. S., Paxton, R. J., Palla, S. L., Yang, G., Askins, M. A., Joy, S. E., et al. (2012). Feasibility, reliability, and validity of the Pediatric Quality of Life Inventory™ Generic Core Scales, Cancer Module, and Multidimensional Fatigue Scale in long-term adult survivors of pediatric cancer. *Pediatric Blood & Cancer, 59*, 703–707.

42. Gordijn, M. S., Cremers, E. M., Kaspers, G. J., & Gemke, R. J. (2011). Fatigue in children: Reliability and validity of the Dutch PedsQL™ Multidimensional Fatigue Scale. *Quality of Life Research, 20*, 1103–1108.

43. Kobayashi, K., Okano, Y., & Hohashi, N. (2011). Reliability and validity of the PedsQL™ Multidimensional Fatigue Scale in Japan. *Quality of Life Research, 20*, 1091–1102.

44. Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430–457.

45. Aday, L. A. (1996). *Designing and conducting health surveys: A comprehensive guide* (2nd ed.). San Francisco: Jossey-Bass.

46. Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Thousand Oaks, CA: Sage.

47. Schwarz, N., & Sudman, N. (Eds.). (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass.

48. Ringold, S., Wallace, C. A., & Rivara, F. P. (2009). Health-related quality of life, physical function, fatigue, and disease activity in children with established polyarticular juvenile idiopathic arthritis. *Journal of Rheumatology, 36*, 1330–1336.

49. Dampier, C., Lieff, S., Lebeau, P., Rhee, S., McMurray, M., Rogers, Z., et al. (2010). Health-related quality of life in children with sickle cell disease: A report from the Comprehensive Sickle Cell Centers Clinical Trial Consortium. *Pediatric Blood & Cancer, 55*, 485–494.

50. Fairclough, D. L. (2002). *Design and analysis of quality of life studies in clinical trials: Interdisciplinary statistics*. New York: Chapman & Hall/CRC.

51. Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

52. Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*, 189–225.

53. Reise, S., Morizot, J., & Hays, R. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19–31.

54. Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*, 544–559.

55. Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling, 13*(2), 287–321.

56. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255.

57. Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568–592.

58. Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246.

59. McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6*(1), 97–103.

60. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

61. Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

62. Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

63. McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6*, 97–103.

64. Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

65. Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275–340). Mahwah, NJ: Lawrence Erlbaum.

66. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

67. Sivo, S. A., Xitao, F., Witta, E. L., & Willse, J. T. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education, 74*(3), 267–288.

68. Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Unpublished Doctoral Dissertation, University of California, Los Angeles. Available from ProQuest Dissertations and Theses (AAT 3066425).

69. Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320–341.

70. Team, R. D. C. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

71. Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36.

72. Revelle, W. (2012). Psych: Procedures for psychological, psychometric, and personality research (Version 1.2.4) [Computer software]. Evanston, IL: Northwestern University.

73. Green, S. B., Akey, T. M., Fleming, K. K., Hershberger, S. L., & Marquis, J. G. (1997). Effect of the number of scale points on Chi square fit indices in confirmatory factor analysis. *Structural Equation Modeling, 4*, 108–120.

74. West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. A. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56–75). Newbury Park, CA: Sage.

75. Bentler, P. M., & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*, 78–117.

76. Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519–530.

77. Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.

78. Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas Jackson at seventy* (pp. 41–71). Boston: Kluwer.

79. Osborne, J. W., Costello, A. B., & Kellow, T. J. (2008). Best practices in exploratory factor analysis. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 86–99). Los Angeles: Sage.

80. Garrido, L. E., Abad, F. J., & Ponsoda, V. (2011). Performance of Velicer's minimum average partial factor retention method with

categorical variables. *Educational and Psychological Measurement, 71*, 551–570.

81. Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousands Oaks, CA: Sage.

82. Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16–29.

83. Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods, 6*, 352–370.

84. Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and Mcdonald's ωh: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*, 123–133.

85. Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*, 517–529.

86. American Educational Research Association, A. P. A., & National Council on Measurement in Education. (1999). American Educational Research Association, American Psychological Association, & National Council on Measurement in Education Standards for educational and psychological testing. Washington, DC.

87. ter Wolbeek, M., van Doornen, L. J. P., Kavelaars, A., & Heijnen, C. J. (2006). Severe fatigue in adolescents: A common phenomenon? *Pediatrics, 117*, e1078–e1086.

88. Viner, R. M., Clark, C., Taylor, S. J. C., Bhui, K., Klineberg, E., Head, J., et al. (2008). Longitudinal risk factors for persistent fatigue in adolescents. *Archives of Pediatrics and Adolescent Medicine, 162*, 469–475.

89. ter Wolbeek, M., van Doornen, L. J. P., Kavelaars, A., & Heijnen, C. J. (2008). Predictors of persistent and new-onset fatigue in adolescent girls. *Pediatrics, 121*, e449–e457.

90. Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*, 129–140.

91. Satorra, A., & Bentler, P. M. (2001). A scaled difference Chi square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514.