

Factorial invariance of child self-report across socioeconomic status groups: a multigroup confirmatory factor analysis utilizing the PedsQL™ 4.0 Generic Core Scales

Christine A. Limbers · Daniel A. Newman ·
James W. Varni

Accepted: June 20, 2008 / Published online: July 8, 2008
© Springer Science+Business Media, LLC 2008

Abstract The measurement of health-related quality of life (HRQOL) in pediatric medicine and health services research has grown significantly over the past decade. Socioeconomic status (SES) differences in health outcomes have been extensively documented in adult and child populations. In order for HRQOL comparisons to be meaningful across different socioeconomic status (SES) groups, items on a HRQOL measure must have equivalent meaning across the SES subpopulations being compared. That is, they must demonstrate factorial invariance. This study examined factorial invariance of child self-report for ages 5–18 across SES groups in 453 children utilizing the PedsQL™ 4.0 Generic Core Scales. Multigroup Confirmatory Factor Analysis was performed specifying a five-factor model across two SES groups. SES groupings were assigned according to the Hollingshead Index of Social Status. Factorial invariance across socioeconomic status groups was demonstrated based on stability of the Comparative Fit Index (CFI) between the models, and several

additional indices of practical fit including the Root Mean Squared Error of Approximation (RMSEA), the Non-Normed Fit Index (NNFI), and the Parsimony Normed Fit Index (PNFI). The findings support an equivalent five-factor structure of child self-report on the PedsQL™ across the two SES groups studied. Based on these data, it can be concluded that children across SES groups interpreted items on the PedsQL™ in a similar manner.

Keywords PedsQL™ · Factorial invariance · Socioeconomic status · Child self-report · Confirmatory factor analysis · Health-related quality of life

Introduction

The measurement of pediatric health-related quality of life (HRQOL) in pediatric medicine and health services research has grown significantly over the past decade (Matza et al. 2004). HRQOL instruments are multidimensional, consisting at the minimum of the physical, psychological (including emotional and cognitive), and social health dimensions delineated by the World Health Organization (FDA 2006; World Health Organization 1948). Although the measurement of pediatric self-reported HRQOL in clinical trials and school and community populations has been advocated for a number of years (Eiser 2004; Varni et al. 2003b), the emerging paradigm shift toward patient-reported outcomes has provided the opportunity to further emphasize the value and critical need for pediatric patient self-report measurement as efficacy outcomes in clinical trials (Connelly and Rapoff 2006; FDA 2006; Razzouk et al. 2006; Schwimmer et al. 2005) and population-based investigations (Seid et al. 2006).

C. A. Limbers
Department of Psychology, College of Liberal Arts, Texas A&M
University, College Station, TX, USA

D. A. Newman
Department of Psychology, University of Illinois at Urbana-
Champaign, 603 East Daniel St., Champaign, IL, USA

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@archmail.tamu.edu
URL: <http://www.pedsqol.org>

By definition, patient-reported outcomes are self-report instruments that directly measure the patient's perceptions of the impact of disease and treatment as clinical trial endpoints (FDA 2006). Patient-reported outcomes include multi-item health-related quality of life instruments, as well as single-item symptom measures (e.g., pain visual analogue scale [VAS]) (Acquadro et al. 2003; Sherman et al. 2006; Willke et al. 2004). Recently, the Food and Drug Administration was quite definitive in their support of patient-reported outcomes by stating that "some treatment effects are known only to the patient" (FDA 2006). However, what has been an obvious recognition in clinical trials for adult patients, that is, that patient-reported outcomes are *patient self-reported* outcomes, has not received the same level of recognition in clinical trials for pediatric patients (Clarke and Eiser 2004). Consequently, demonstrating the measurement properties of child self-reported HRQOL instruments is an essential step toward meeting the requirements by the Food and Drug Administration for patient-reported outcomes in pediatric clinical trials, as well as other comparative group analyses.

Socioeconomic status differences and health-related quality of life measurement

Socioeconomic status (SES) differences in health outcomes have been extensively documented in adult and child populations (Chen et al. 2002; Denvir et al. 2006; Montazeri et al. 2003; Olson et al. 2004). A comprehensive review demonstrated a SES—health gradient in pediatric populations, with decreases in SES found to be associated in a monotonic fashion with increases in health problems (Chen et al. 2002). While reviews of the literature point to an increasingly clear picture of the effects of SES on child health outcomes, less evidence has been documented on the effects of SES on the broader concept of pediatric HRQOL (Bradley and Corwyn 2002). In children and adolescents, SES differences in self-reported HRQOL have been demonstrated irrespective of the instrument utilized (Varni et al. 2006b; von Rueden et al. 2006). For example, children in low SES schools reported significantly lower PedsQL™ Total Scale scores than children in low-middle SES schools and middle SES schools (Varni et al. 2006b).

Cultural and social differences among SES groups may lead to differences in the conceptualization of health-related quality of life. For example, some research with adults suggests that lower SES groups have the propensity to incorporate physical symptoms into their responses to questions designed to evaluate emotional functioning (Nguyen et al. 2004). Thus, items on a HRQOL measure that assess emotional and physical functioning may be understood differently by respondents in different SES groups. Specifically, for lower SES groups, the prevalence

of physical symptoms may be a stronger marker of impairments in emotional functioning than affective symptoms (Nguyen et al. 2004). To date, no published studies have examined the factor structure of the PedsQL™ 4.0 Generic Core Scales or other pediatric health-related quality of life measures across different socioeconomic status groups. In order to have greater confidence that a HRQOL instrument is measuring the same constructs across different SES groups (i.e., that the items have the same meaning for all participants regardless of their SES), it is essential to demonstrate factorial invariance (Gregorich 2006; Meredith and Teresi 2006). This is particularly important in SES investigations given the educational achievement differences inherent across SES gradients for parents and their children.

Factorial invariance

Generic HRQOL instruments facilitate comparisons across diverse chronic health conditions, as well as benchmarking with healthy and community populations (Patrick and Deyo 1989; Varni et al. 1999). However, in order for these comparisons to be valid, items on a HRQOL measure must have equivalent meaning across the subpopulations being compared (French and Finch 2006). That is, they must demonstrate factorial invariance (Cheung and Rensvold 2002; Vandenberg and Lance 2000). Horn and McArdle (Horn and McArdle 1992) defined two hierarchical levels of factorial invariance. The first level of factorial invariance, which necessitates that the psychometric properties of an instrument are equal across groups (i.e., configural and metric invariance), must be established prior to testing a higher level of factorial invariance, which requires equality of item-specific intercepts across groups (i.e., scalar invariance) (French and Finch 2006; Horn and McArdle 1992; Little 1997). Configural invariance is a test in which the same pattern of fixed and free factor loadings is specified for each group (Vandenberg and Lance 2000), while metric invariance is a test where factor loadings for like items are held invariant across groups (Vandenberg and Lance 2000). Scalar invariance assesses whether intercepts of like items' regressions on the latent variables are invariant across groups (Vandenberg and Lance 2000). Multigroup Confirmatory Factor Analysis (CFA) is a method used to assess these different levels of factorial invariance (French and Finch 2006; Vandenberg and Lance 2000). To the degree that components of the factor model are determined to be equal across subpopulations, factorial invariance of an instrument can be inferred (French and Finch 2006).

While the use of Multigroup CFA for invariance testing has grown considerably in recent years (French and Finch 2006; Vandenberg and Lance 2000), there has been a rel-

ative absence of studies that have examined the factorial invariance of pediatric HRQOL measures across socioeconomic status groups. Furthermore, the majority of studies to date have been concerned with establishing configural and metric invariance, ignoring higher levels of factorial invariance (Gregorich 2006; Vandenberg and Lance 2000). Without establishing stricter levels of factorial invariance such as scalar invariance, differences in observed health outcomes across socioeconomic status groups may be confounded by differences in item-specific intercepts (Meredith and Teresi 2006).

The 23-item PedsQL™ 4.0 Generic Core Scales is a health-related quality of life measure in which feasibility, reliability and validity have been supported for pediatric patients with chronic health conditions and healthy school and community populations (Varni et al. 2006a; 2003a, 2002a, 2004, 2006b, 2003b, 2007a, 2007b, 2002b, 2001), with international data on tens of thousands of healthy children and children with pediatric chronic health conditions published or in press in over 230 peer-reviewed journal articles since the 2001 publication of the PedsQL™ 4.0 Generic Core Scales in the journal *Medical Care* (see Publications, www.pedsq.org). The PedsQL™ has been translated into over 60 international languages for both research and clinical practice including, for example, German (Felder-Puig et al. in press), Catalan (Huguet and Miro 2008), Turkish (Laaksonen et al. 2007), Lithuanian (Makari et al. 2007), and Dutch (Reinfjell et al. 2006). The PedsQL™ 4.0 has demonstrated feasibility across numerous studies, as evidenced by requiring less than 5 min to complete and averaging less than 1% missing responses for the Total Scale Score. The objective of the present study was to examine the factorial invariance of children's self-reported HRQOL on the PedsQL™ 4.0 Generic Core Scales across socioeconomic status groups using the Multigroup CFA framework (Vandenberg and Lance 2000).

Methods

Participants and settings

The sample contains child self-report data on 453 children ages 5–18 years from the PedsQL™ 4.0 Generic Core Scales database (previously published data, $n = 453$, 100.0%). Participants were recruited from subspecialty clinics and hospitals in which they were being seen for chronic illness care (cancer, $n = 128$, 28.3%; asthma, $n = 13$, 2.9%; cardiac conditions, $n = 90$, 19.9%; rheumatic condition, $n = 60$, 13.2%; diabetes, $n = 110$, 24.3%; cerebral palsy, $n = 31$, 6.8%; end stage renal disease, $n = 21$, 4.6%). Participants were assessed in-person

($n = 446$, 98.5%) or by telephone ($n = 7$, 1.5%). For in-person mode of administration, research assistants obtained written parental informed consent and child assent. Paper-and-pencil surveys were self-administered for parents and for children ages 8–18, and interview-administered for children ages 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 ages 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™ 4.0 individually to the parent and 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 the Institutional Review Board at Children's Hospital and Health Center, San Diego and other appropriate local Institutional Review Boards and were carried out in accordance with universal ethical principles.

Measures

The PedsQL™ 4.0 (Pediatric Quality of Life Inventory™ Version 4.0)

The a priori conceptually-derived measurement model for the 23-item PedsQL™ 4.0 Generic Core Scales specified four scales: (1) Physical Functioning (8 items), (2) Emotional Functioning (5 items), (3) Social Functioning (5 items), and (4) School Functioning (5 items), which were developed through focus groups, cognitive interviews, pre-testing, and field testing measurement development protocols (Varni et al. 2001, 1999).

An empirically-derived measurement model from the initial field test of the PedsQL™ 4.0 Generic Core Scales utilizing principal components analysis with oblique rotation on the 23 items to test the underlying dimensions indicated that a five-factor model is appropriate (Varni et al. 2001). Specifically, the School Functioning Scale items split into two different factors, with three items measuring school-related cognitive functioning and two items measuring days missed from school due to illness. Thus, based on these findings from the initial field test, a five-factor model was specified for the CFA analyses in the current study.

The PedsQL™ 4.0 Generic Core Scales are comprised of parallel child self-report and parent proxy-report formats. Child self-report includes ages 5–7, 8–12, and 13–18 years. The items for each of the forms are essentially identical, differing in developmentally appropriate language. The instructions ask how much of a problem each item has been during the past one month. A 5-point cate-

gorical response scale is utilized across child self-report for ages 8–18 (0 = never a problem; 1 = almost never a problem; 2 = sometimes a problem; 3 = often a problem; 4 = almost always a problem). 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. Scale scores are computed as the sum of the items divided by the number of items answered. To further increase the ease of use for the young child self-report (ages 5–7), the response scale is reworded and simplified to a 3-point scale (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. For the purposes of these analyses, child self-report data across socioeconomic status groups for children 5–18 years were utilized.

PedsQL™ Family Information Form

The PedsQL™ Family Information Form, completed by parents (Varni et al. 2001), contains demographic information including the child's date of birth, gender, race/ethnicity, and parental education and occupation information.

For these analyses, socioeconomic status groupings were assigned according to the Hollingshead Index of Social Status (Hollingshead 1975), which was calculated for each child based on the parental education and occupation reported by parents on the PedsQL™ Family Information Form. The Hollingshead Index integrates an ordinal ranking of nine occupational categories, ranging from higher executives, proprietors of large businesses, and major professionals (9) to farm laborers/mental service workers (1), with a seven category ranking of educational categories ranging from attainment of a professional degree (7) to completing less than the seventh grade (1) (Hollingshead 1975). The occupational score is weighted by five and the educational score is weighted by 3; these two weighted scores are added together to create an overall score (Hollingshead 1975). If both parents are employed, the education and occupation scores for the father and mother are summed and divided by two (Hollingshead 1975). If only one parent is employed, the education and occupation scores for that parent are utilized. Overall scores range from a high of 66 to a low of 8. Scores ranging from 66 to 55 are designated as upper social status, 54–40 are designated as upper-middle social status, 39–30 are designated as middle social status, 29–20 are designated as lower-middle social status, and 19–8 are designated as lower social status (Hollingshead 1975).

For the present study, the upper (66–55), upper-middle (54–40), and middle (30–39) social status groups designated

by Hollingshead (Hollingshead 1975) were combined into one group (Upper), as were the lower-middle (29–20) and lower (19–8) social status groups (Lower). Grouping participants in this manner initially resulted in a total of 828 participants in the Upper SES group and 259 participants in the Lower SES group. Given the strong association between race/ethnicity and socioeconomic status (Barrett et al. 1995), in order to control for race/ethnicity in our analyses, we randomly matched the Upper socioeconomic status group on race/ethnicity to the Lower socioeconomic status group utilizing the SPSS Version 15.0 statistical software random sample case selection command (SPSS 2007). Due to the disproportionate number of White-Non-Hispanic participants in the Upper SES group (59%), this resulted in excluding 616 participants in the Upper SES group from the analyses. Thus, the final sample for the Upper SES group (Hollingshead Index ranged from 30 to 66) randomly matched by race/ethnicity to the Lower SES group was 212. The final sample for the Lower SES group (Hollingshead Index ranged from 8 to 29) after excluding participants that had missing race/ethnicity data was 241. As such, for the present analyses, the heuristic recommendations set forth by Guilford and Kline, which suggest a minimum sample size of 200 per group when conducting confirmatory factor analysis (Guilford 1954; Kline 2005; MacCallum et al. 2001), were met.

Statistical analyses

A five-factor model was specified for the CFA analyses. We first conducted an omnibus test of the equality of sample covariance matrices (Model 0) (Cheung and Rensvold 2002; Vandenberg and Lance 2000). The omnibus test of the equality of covariance matrices simultaneously compares the groups on all parameters in the covariance matrix (O'Brien 1992). Thus, failure to reject the null hypothesis that the covariance matrices are invariant across groups is an indication of overall measurement equivalence across groups and consequently this test serves as an important first step in assessing factorial invariance (Vandenberg and Lance 2000). We next assessed configural invariance (Model 1). For the configural invariance model, the pattern of fixed and free factor loadings was held constant across the socioeconomic status groups while the magnitudes of these loadings were not constrained to be equal (Vandenberg and Lance 2000). Subsequently, metric invariance (Model 2) was tested by constraining corresponding factor loadings to be equal across groups (Horn and McArdle 1992; Vandenberg and Lance 2000). We proceeded by testing the scalar invariance (strong factorial invariance) model. This model (Model 3) further imposed equality constraints on the item

intercepts across socioeconomic status groups (Vandenberg and Lance 2000). Following Cheung and Rensvold’s (Cheung and Rensvold 1999) recommendation, we fixed one item loading at unity for each factor, and also allowed the five latent factors of the PedsQL™ 4.0 Generic Core Scales to intercorrelate.

The chi-square statistic is an extremely sensitive statistical test, and with large samples such as the present one ($n = 453$), is not a practical test of model fit (Cheung and Rensvold 2002; Maydeu-Olivares and D’Zurilla 1996). Thus, consistent with recommendations set forth in the literature (Bentler 1990; Cheung and Rensvold 2002; Mulaik et al. 1989), we examined several different model indices of practical fit including the Root Mean Squared Error of Approximation (RMSEA) (Steiger 1990), the Comparative Fit Index (CFI) (Bentler 1990), the Tucker-Lewis Index (also known as the Non-Normed Fit Index; NNFI) (Tucker and Lewis 1973), and the Parsimony Normed Fit Index (PNFI) (James et al. 1982). For the RMSEA, values in the range of 0.05–0.08 were taken to indicate acceptable fit, values in the range of 0.08–0.10 to indicate marginal fit, and values larger than 0.10 to indicate poor fit (Browne and Cudeck 1992; Steiger 1989). For the NNFI and CFI indices, acceptable model fit was suggested by values greater than or equal to 0.90 (Bentler 1990; Hu and Bentler 1995; Mulaik et al. 1989). For parsimony indices however, such as the PNFI, the concept of optimal values may not pertain given that parsimonious indices are mainly influenced by model degrees of freedom (Sivo et al. 2006). We therefore do not invoke any absolute standard of model fit for the PNFI, but rather note simply that higher PNFI values reflect more parsimonious fit (Sivo et al. 2006).

For evaluating the relative fit of nested measurement equivalence models, Cheung and Rensvold (Cheung and Rensvold 2002) recommend that an appropriate criterion is the change in CFI, because large sample sizes, such as ours, will almost always signal a statistically significant chi-square difference, even when the model parameter differences are trivial (Cheung and Rensvold 2002; Vandenberg and Lance 2000). When conducting Multigroup CFA to assess factorial invariance, it is recommended that changes in CFI values of -0.01 or less are indicative of factorial invariance across groups (Cheung and Rensvold 2002; Vandenberg and Lance 2000). While it has been proposed that changes in CFI values from -0.01 to -0.02 may also be indicative of factorial invariance across groups (Vandenberg and Lance 2000), for the present analyses changes in CFI values of -0.01 or less were taken to indicate factorial invariance across groups (Cheung and Rensvold 2002; Vandenberg and Lance 2000). In each model comparison, the change in CFI was used to test whether the more constrained model (i.e., the model imposing more

equality constraints across groups) resulted in worse fit. Statistical analyses were conducted using LISREL 8.5 for Windows (Joreskog and Sorbom 2003).

Results

Participant characteristics

Table 1 presents the age, gender, and race/ethnicity of the participants across the socioeconomic status groups, and suggests that with the exception of gender ($P = 0.04$), these characteristics do not differ significantly between the socioeconomic status groups in the current sample.

Factorial invariance: goodness of fit indexes

Table 2 presents the chi-square values (with df), fit indices, and change in CFI values for the measurement invariance models. Model 0 (invariant covariance matrices) demonstrates acceptable model fit (RMSEA = 0.0694; CFI = 0.956; NNFI = 0.920). Taken together, the fit indices for Model 1 (*configural invariance*) indicate that the five-factor measurement model of the PedsQL™ 4.0 Generic Core Scales has acceptable fit within each socioeconomic status group (RMSEA = 0.0867; CFI = 0.920; NNFI = 0.908). Based on the change in CFI criterion recommended by Cheung and Rensvold (Cheung and Rensvold 2002) and the indices of practical fit taken together, we conclude that Model 2 (the *metric invariance* model) fits as well as Model 1, supporting metric invariance

Table 1 Characteristics of Study Participants across Socioeconomic Status Groups

	Lower SES ($n = 241$)	Upper SES ($n = 212$)	<i>P</i> -value
Characteristics			
Age (years)			
Mean (SD)	11.79 (3.86)	12.53 (3.75)	0.05 ^a
Gender			
Boys (%)	128 (53%)	92 (43%)	0.04*
Girls (%)	113 (47%)	120 (57%)	
Race/Ethnicity			
White, Non-Hispanic (%)	77 (32%)	74 (35%)	0.99**
Hispanic (%)	121 (50%)	103 (49%)	
Black, Non-Hispanic (%)	21 (9%)	17 (8%)	
Asian/Pacific Islander (%)	7 (3%)	5 (2%)	
American Indian/ Alaskan Native (%)	11 (5%)	9 (4%)	
Other (%)	4 (2%)	4(2%)	

Note: ^a Independent Samples *t*-test; * Chi square test $df = 1$; ** Chi square test $df = 5$

Table 2 Health-related quality of life and socioeconomic status: χ^2 Values (with *df*), goodness of fit indices, and change in CFI values for measurement invariance models

Model of invariance (across socioeconomic status groups)	df	χ^2	RMSEA	CFI	NNFI	PNFI	Δ CFI
0. Invariant covariance matrices	276	575	0.0694	0.956	0.920	0.506	–
1. Configural invariance	440	1184	0.0867	0.920	0.908	0.762	–
2. Metric invariance	458	1212	0.0856	0.919	0.910	0.790	–0.001
3. Scalar invariance	476	1234	0.0841	0.918	0.913	0.819	–0.001

(RMSEA = 0.0856; CFI = 0.919; NNFI = 0.910; Δ CFI = –0.001). We conclude that the *scalar invariance* model (Model 3) fits nearly as well as Model 2 (RMSEA = 0.0841; CFI = 0.918; NNFI = 0.913; Δ CFI = –0.001), supporting that item intercepts are relatively invariant across socioeconomic status groups (Vandenberg and Lance 2000). Considering all practical fit indices together we judge the five-factor measurement model of the PedsQL™ 4.0 Generic Core Scales to have acceptable fit across the socioeconomic status groups.

PNFI values (Table 2) give an advantage to the model imposing the most equality constraints across the socioeconomic status groups (Model 3), because it estimates fewer parameters and is thus more parsimonious (Mulaik et al. 1989). In short, estimating 2 separate measurement models across the 2 socioeconomic status groups (Model 1) results in similar empirical fit as estimating a single measurement model that constrains the factor loadings and item intercepts to be equal across socioeconomic status groups (Model 3; overall change in CFI from Model 1 to Model 3 is –0.002, from 0.920 to 0.918). However, using the model imposing the most equality constraints (Model 3) is much more parsimonious (PNFI improves from 0.762 to 0.819).

Table 3 presents the factor loadings for two models: (a) Model 1 (configural invariance model, in which factor loadings are estimated separately within each socioeconomic status group), and (b) Model 2 (metric invariance model, in which a single set of factor loadings is estimated for the socioeconomic status groups). Estimates in Table 3 reflect the common metric completely standardized solution (for Model 1 and Model 2). To be able to compare factor parameters across samples when conducting a Multigroup CFA, the variables in each group must be measured in a common metric (Joreskog and Sorbom 1996). In Lisrel, the common metric completely standardized solution standardizes the observed variables such that factor loadings can be compared across groups on a common scale (Joreskog and Sorbom 1996). Overall across the socioeconomic status groups, the items on the PedsQL™ 4.0 Generic Core Scales load similarly onto their 5 corresponding factors, providing further evidence that the items are measuring similar HRQOL constructs across socioeconomic status groups. It could be noted post hoc,

however, that the factor loadings for three of the twenty-three items indicate that these items may not be operating identically across the Lower and Upper socioeconomic status groups. For these three items (Physical 7—I hurt or ache; Emotional 3—I feel angry; and Emotional 4—I have trouble sleeping), there is a general trend for the items to have higher factor loadings for the Upper socioeconomic status group.

Intercorrelations among the five PedsQL™ latent factors (averaged across groups) are in the range from 0.448 to 0.777. Intercorrelations among the five PedsQL™ latent factors for the Lower socioeconomic status group range from 0.423 to 0.833, while for the Upper socioeconomic status group they range from 0.272 to 0.710. Although these are relatively large factor intercorrelations, we do not judge them to be so large as to threaten the discriminant validity of the PedsQL™ 4.0 Generic Core Scales. That is, factor correlations significantly smaller than 1.0 suggest discriminant validity (Campbell and Fiske 1959; Widaman 1985). Although, the five PedsQL™ factors are intercorrelated (oblique), they are similarly intercorrelated across socioeconomic status groups, and therefore the factor obliqueness does not threaten the measurement equivalence of the scales.

Discussion

The present findings demonstrate that when self-reporting their HRQOL, children across the two socioeconomic status groups studied had a similar five-factor HRQOL structure on the PedsQL™ 4.0 Generic Core Scales. As a result, it can be concluded that children in this study had a similar interpretation of the items on the PedsQL™ 4.0 Generic Core Scales regardless of their socioeconomic status.

These findings contribute to the empirical literature on the PedsQL™ Measurement Model by demonstrating factorial invariance for child self-report across two SES groups. These findings complement recent findings that the PedsQL™ 4.0 Generic Core Scales demonstrate factorial invariance for child self-report across 12 individual age groups (Limbers et al. in press) and across health status

Table 3 Health-related quality of life and socioeconomic status: standardized factor loadings for the configural and metric invariance models

Item	Model 1		Model 2
	Lower SES	Upper SES	
Phys1. It is hard for me to walk more than one block	0.75	0.65	0.70
Phys2. It is hard for me to run	0.72	0.72	0.72
Phys3. It is hard for me to do sports activity or exercise	0.82	0.75	0.79
Phys4. It is hard for me to lift something heavy	0.56	0.67	0.61
Phys5. It is hard for me to take a bath or shower by myself	0.55	0.49	0.52
Phys6. It is hard for me to do chores around the house	0.57	0.52	0.54
Phys7. I hurt or ache	0.40	0.68	0.54
Phys8. I have low energy	0.56	0.70	0.64
Emot1. I feel afraid or scared	0.69	0.63	0.65
Emot2. I feel sad or blue	0.71	0.71	0.69
Emot3. I feel angry	0.52	0.77	0.65
Emot4. I have trouble sleeping	0.34	0.51	0.43
Emot5. I worry about what will happen to me	0.55	0.63	0.59
Soc1. I have trouble getting along with other kids	0.50	0.36	0.43
Soc2. Other kids do not want to be my friend	0.33	0.39	0.37
Soc3. Other kids tease me	0.42	0.41	0.42
Soc4. I cannot do things that other kids my age can do	0.61	0.57	0.59
Soc5. It is hard to keep up when I play with other kids	0.73	0.73	0.73
Sch1. It is hard to pay attention in class	0.63	0.61	0.62
Sch2. I forget things	0.77	0.64	0.71
Sch3. I have trouble keeping up with my schoolwork	0.72	0.72	0.73
MsSch1. I miss school because of not feeling well	0.81	0.75	0.78
MsSch2. I miss school to go to the doctor or hospital	0.56	0.56	0.56

Note: Model 1 = configural invariance model. Model 2 = metric invariance model

Phys = Physical functioning, Emot = Emotional functioning, Soc = Social functioning

Sch = School functioning, MsSch = Missed school

(Limbers et al. 2008). As the consumers of pediatric healthcare, children and adolescents are uniquely positioned to give their self-reported perspectives on healthcare quality through perceptions of their own health-related quality of life outcomes.

The statistical methods utilized in the present study have important implications in general for research studies in which different SES groups are recruited, and specifically for studies in which the HRQOL of different SES groups are analyzed with the PedsQL™ 4.0 Generic Core Scales (Hassan et al. 2006). The results of the present study suggest that when differences are found across groups when utilizing the PedsQL™, these differences are more likely real differences in self-perceived HRQOL for the groups studied, rather than differences in interpretation of the PedsQL™ items as a function of SES.

Standardized HRQOL instruments must demonstrate that test items are interpreted similarly across age, gender, language, socioeconomic, and race/ethnicity subpopulations (Gregorich 2006; Meredith and Teresi 2006), as well

as mode of administration and health status categories. Demonstrating stricter levels of the hierarchy of factorial invariance across these subpopulations and methods is critical given the reliance of clinical trials, health disparities and comparative public health research on self-report instruments (Gregorich 2006; Meredith and Teresi 2006). Gregorich recently emphasized that there has been a general lack of comparative studies in medical and health services research utilizing self-report measures that have adequately demonstrated stricter levels of factorial invariance (Gregorich 2006). Further, he proposed that self-report instruments demonstrating stricter levels in the hierarchy of factorial invariance in some population groups should also be shown to demonstrate factorial invariance in additional, relevant populations. Thus, the establishment of factorial invariance is an iterative process, and must be replicated periodically as, for example, language groups, cultural norms, methods of administration, and the compositions of race/ethnicity and socioeconomic subpopulations change over time (Gregorich 2006).

The present study has several limitations. First, the socioeconomic status groups in our study differed significantly with regard to gender. Specifically, there were a higher percentage of girls in the Upper socioeconomic status group compared to the Lower socioeconomic status group (Table 1). To the best of our knowledge, no previously published studies have investigated the factor stability of the PedsQL™ 4.0 Generic Core Scale child self-report across gender. However, previous research with healthy children and children with chronic health conditions utilizing the PedsQL™ 4.0 Generic Core Scales child self-report has demonstrated no significant differences in PedsQL™ mean scores across gender (Berrin et al. 2007; Stern et al. 2007; Uzark et al. 2008; Varni et al. 2001). Further, it should be noted that any similarities in the factor structures of PedsQL™ 4.0 Generic Core Scales across the socioeconomic status subpopulations in the present study reflect similarities in the measurement properties of the scale *despite* gender differences across socioeconomic status groups. That is, our test of measurement invariance across socioeconomic status subpopulations represents a conservative test of measurement equivalence, since gender differences across socioeconomic status groups would be expected to exacerbate, not attenuate, groupwise differences in factor structures.

The large percentage of Hispanic participants in both SES groups may also limit the generalizability of these findings. In addition, for three items (Physical 7—I hurt or ache; Emotional 3—I feel angry; and Emotional 4—I have trouble sleeping), there is a general trend for the items to have higher factor loadings for the Upper socioeconomic status group. Thus, despite the acceptable fit of the five-factor model across socioeconomic status groups (i.e., both SES groups demonstrated equivalence in the pattern of factor structure suggesting that the relationship between the factors and the variables by which the factors are being measured is reliable), the construct validity for some items may be slightly better for respondents in the Upper socioeconomic status group. The response thresholds for reporting pain, sleeping problems, and anger may be higher in lower socioeconomic status groups, as the privilege to express these symptoms could be less normative for lower socioeconomic status children. Further research on the differential item characteristics curves of these three items (within an Item Response Theory framework) may be warranted. If this finding—that these three HRQOL items show lower loadings for low SES groups—replicates in future work, then there may be grounds for item revision or rescaling when the PedsQL™ is used to compare high versus low SES respondents. In the meantime, we mention that future SES comparative research on HRQOL be conducted with the awareness that inclusion of these three items may lead to subtle differences across groups in the

underlying constructs being compared. If a researcher is especially concerned about these loading differences, then SES groups could be compared using stripped-down subscales from which these items have been removed, as an addition/supplement to SES comparisons on the full, validated scale.

Finally, there are multiple methods available for measuring SES, each with documented strengths and limitations (Cirino et al. 2002). The Hollingshead Index of Social Status was utilized in the present study because it is one of the most commonly used SES measure in the US, and has demonstrated a high level of agreement with other commonly used SES measures (Cirino et al. 2002). For the present study, larger sample sizes would have provided an opportunity for a more fine grained analysis of Hollingshead's five SES groups. In the present study we combined the upper (66–55), upper-middle (54–40), and middle (30–39) social status groups designated by Hollingshead (Hollingshead 1975) into one group (Upper), and the lower-middle (29–20) and lower social status (19–8) groups designated by Hollingshead (Lower) into another group. We then randomly matched the Upper SES group to the Lower SES group on race/ethnicity given the strong association between race/ethnicity and socioeconomic status (Barrett et al. 1995). Doing so required us to exclude a large number of cases in the Upper SES group ($n = 616$) which has the potential to bias parameter estimates. It should be noted, however, that when examining factorial invariance of the PedsQL™ 4.0 Generic Core Scales across SES groups for child self-report in the full sample of 1,087 children, not matched on race/ethnicity, scalar invariance across socioeconomic status groups was demonstrated based on stability of the CFI between the models (Δ CFI values were all below -0.01 ; CFI's were all at or above 0.936) and additional indices of practical fit including the RMSEA and the NNFI (RMSEA's were all at or below 0.077; NNFI's were all at or above 0.931). Taken as a whole, these data support the robustness of our findings of an equivalent five-factor structure of child self-report on the PedsQL™ across the SES groups studied.

In sum, quantitative comparative research requires that instruments measure constructs with the same meaning across groups, which has been inadequately reported in the extant comparative literature using self-report instruments (Gregorich 2006). Self-reported clinical outcomes such as HRQOL that are not directly observed require latent variable models such as Multigroup CFA to demonstrate factorial invariance across groups to support valid assessment (Meredith and Teresi 2006). The PedsQL™ 4.0 Generic Core Scales appear to be interpreted equivalently, and thus to measure the same constructs, irrespective of the socioeconomic status group of the children who provide their self-reported HRQOL.

Acknowledgement Preparation of this manuscript was supported by an intramural grant from the Texas A&M University Research Foundation.

Competing Interests: Dr. Varni holds the copyright and the trademark for the PedsQL™ and receives financial compensation from the Mapi Research Trust, which is a nonprofit research institute that charges distribution fees to for-profit companies that use the Pediatric Quality of Life Inventory™.

References

- Acquadro, C., Berzon, R., Dubois, D., Leidy, N. K., Marquis, P., Revicki, D., et al. (2003). Incorporating the patient's perspective into drug development and communication: An ad hoc task force report of the patient-reported outcomes (PRO) harmonization group meeting at the Food and Drug Administration, February 16, 2001. *Value in Health*, 6, 522–531. doi:10.1046/j.1524-4733.2003.65309.x.
- Barrett, R. J., Harlan, L. C., Wesley, M. N., Hill, H. A., Chen, V. W., Clayton, L. A., et al. (1995). Endometrial cancer: Stage at diagnosis and associated factors in black and white patients. *American Journal of Obstetrics and Gynecology*, 173, 414–422. doi:10.1016/0002-9378(95)90261-9.
- Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246. doi:10.1037/0033-2909.107.2.238.
- Berrin, S. J., Malcarne, V. L., Varni, J. W., Burwinkle, T., Sherman, S. A., Artavia, K. A., et al. (2007). Pain, fatigue, and school functioning in children with cerebral palsy: A path-analytic model. *Journal of Pediatric Psychology*, 32, 330–337. doi:10.1093/jpepsy/jsl017.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371–399. doi:10.1146/annurev.psych.53.100901.135233.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21, 230–258. doi:10.1177/004912419201002005.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105. doi:10.1037/h0046016.
- Chen, E., Matthews, K. A., & Boyce, W. T. (2002). Socioeconomic differences in children's health: How and why do these relationships change with age? *Psychological Bulletin*, 128, 295–329. doi:10.1037/0033-2909.128.2.295.
- Cheung, G. W., & Rensvold, R. B. (1999). What constitutes significant differences in evaluating measurement invariance? *Conference of the Academy of Management*, Chicago.
- Cheung, G. W., Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255.
- Cirino, P. T., Chin, C. E., Sevcik, R. A., Wolf, M., & Morris, R. D. (2002). Measuring socioeconomic status: Reliability and preliminary validity for different approaches. *Assessment*, 9, 145–155. doi:10.1177/10791102009002005.
- Clarke, S. A., & Eiser, C. (2004). The measurement of health-related quality of life in pediatric clinical trials: A systematic review. *Health and Quality of Life Outcomes*, 2(66), 1–5.
- Connelly, M., & Rapoff, M. A. (2006). Assessing health-related quality of life in children with recurrent headache: Reliability and validity of the PedsQL™ 4.0 in a pediatric sample. *Journal of Pediatric Psychology*, 31, 698–702. doi:10.1093/jpepsy/jsj063
- Denvir, M. A., Lee, A. J., Rysdale, J., Walker, A., Eteiba, H., Starkey, I. R., et al. (2006). Influence of socioeconomic status on clinical outcomes and quality of life after percutaneous coronary intervention. *Journal of Epidemiology and Community Health*, 60, 1085–1088. doi:10.1136/jech.2005.044255.
- Eiser, C. (2004). Use of quality of life measures in clinical trials. *Ambulatory Pediatrics*, 4, 395–399. doi:10.1367/A03-112R.1.
- FDA. (2006). *Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims*. Rockville, MD: Center for Drug Evaluation and Research, Food and Drug Administration.
- Felder-Puig, R., Topf, R., Gadner, H., & Formann, A.K. (in press). Measuring health-related quality of life in children from different perspectives using the Pediatric Quality of Life Inventory™ (PedsQL™) and teachers' ratings. *Journal of Public Health*.
- French, B., & Finch, W. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling*, 13, 378–402. doi:10.1207/s15328007sem1303_3.
- 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. doi:10.1097/01.mlr.0000245454.12228.8f.
- Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.
- Hassan, K., Loar, R., Anderson, B. J., & Heptulla, R. A. (2006). The role of socioeconomic status, depression, quality of life, and glycemic control in Type 1 diabetes mellitus. *The Journal of Pediatrics*, 149, 526–531. doi:10.1016/j.jpeds.2006.05.039.
- Hollingshead, A. B. (1975). *Four factor index of social status*. New Haven, CT: Yale University.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117–144.
- Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications*. Thousand Oaks: Sage.
- Huguet, A., & Miro, J. (2008). Development and psychometric evaluation of a Catalan self- and interviewer-administered version of the Pediatric Quality of Life Inventory™ Version 4.0. *Journal of Pediatric Psychology*, 33, 63–79. doi:10.1093/jpepsy/jsm040.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Models, assumptions, and data*. Beverly Hills: Sage.
- Joreskog, K. G., & Sorbom, D. (1996). *Lisrel 8: User's Reference Guide*. Chicago, IL: Scientific Software International.
- Joreskog, K. G., & Sorbom, D. (2003). *LISREL 8.5*. Lincolnwood, IL: Scientific Software International, Inc.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford.
- Laaksonen, C., Aromaa, M., Heinonen, O. J., Heinonen, S., & Salanterä, S. (2007). Pediatric health-related quality of life instrument for primary school children: Cross-cultural validation. *Journal of Advanced Nursing*, 59, 542–550. doi:10.1111/j.1365-2648.2007.04347.x.
- Limbers, C. A., Newman, D. A., & Varni, J. W. (in press). 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*.
- 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 of ages 5 to 18 years utilizing the PedsQL™ 4.0 Generic Core Scales.

- Journal of Pediatric Psychology*, 33, 630–639. doi:10.1093/jpepsy/jsm131.
- Little, T. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 31, 53–76. doi:10.1207/s15327906mbr3201_3.
- MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, 36, 611–637. doi:10.1207/S15327906MBR3604_06.
- Makari, J., Zaborskis, A., Labanauskas, L., & Rageliene, L. (2007). The quality of life of children suffering from oncologic diseases in Lithuania. *Medicina*, 43, 726–735.
- Matza, L. S., Swensen, A. R., Flood, E. M., Secnik, K., & Leidy, N. K. (2004). Assessment of health-related quality of life in children: A review of conceptual, methodological, and regulatory issues. *Value in Health*, 7, 79–92. doi:10.1111/j.1524-4733.2004.71273.x.
- Maydeu-Olivares, A., & D’Zurilla, T. J. (1996). A factor-analytic study of the Social Problem-Solving Inventory: An integration of theory and data. *Cognitive Therapy and Research*, 20, 115–133. doi:10.1007/BF02228030.
- Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, 44, S69–S77. doi:10.1097/01.mlr.0000245438.73837.89.
- Montazeri, A., Hole, D. J., Milroy, R., McEwen, J., & Gillis, C. R. (2003). Quality of life in lung cancer patients: Does socioeconomic status matter? *Health and Quality of Life Outcomes*, 1, 1–6. doi:10.1186/1477-7525-1-1.
- Mulaik, S., James, L., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105, 430–445. doi:10.1037/0033-2909.105.3.430.
- Nguyen, H. T., Kitner-Triolo, M., Evans, M. K., & Zonderman, A. B. (2004). Factorial invariance of the CES-D in low socioeconomic status African Americans compared with a nationally representative sample. *Psychiatry Research*, 126, 177–187. doi:10.1016/j.psychres.2004.02.004.
- O’Brien, P. C. (1992). Robust procedures for testing equality of covariance matrices. *Biometrics*, 48, 819–827. doi:10.2307/2532347.
- Olson, L. M., Lara, M., & Frintner, M. P. (2004). Measuring health status and quality of life for US children: Relationship to race, ethnicity, and income status. *Ambulatory Pediatrics*, 4, 377–386. doi:10.1367/A03-156.1.
- Patrick, D. L., & Deyo, R. A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Medical Care*, 27, S217–S233. doi:10.1097/00005650-198903001-00018.
- Razzouk, B. I., Hord, J. D., Hockenberry, M., Hinds, P. S., Feusner, J., Williams, D., et al. (2006). Double-blind, placebo-controlled study of quality of life, hematologic end points, and safety of weekly epoetin alfa in children with cancer receiving myelosuppressive chemotherapy. *Journal of Clinical Oncology*, 24, 3583–3589. doi:10.1200/JCO.2005.03.4371.
- 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.
- Schwimmer, J. B., Middleton, M. S., Deutsch, R., & Lavine, J. E. (2005). A phase 2 trial of metformin as a treatment for non-diabetic pediatric non-alcoholic steatohepatitis. *Alimentary Pharmacology & Therapeutics*, 21, 871–879. doi:10.1111/j.1365-2036.2005.02420.x.
- Seid, M., Varni, J. W., Cummings, L., & Schonlau, M. (2006). The impact of realized access to care on health-related quality of life: A two-year prospective cohort study of children in the California State Children’s Health Insurance Program. *The Journal of Pediatrics*, 149, 354–361. doi:10.1016/j.jpeds.2006.04.024.
- Sherman, S. A., Eisen, S., Burwinkle, T. M., & Varni, J. W. (2006). The PedsQL™ present functioning visual analogue scales: Preliminary reliability and validity. *Health and Quality of Life Outcomes*, 4(75), 1–10.
- Sivo, S. A., Fan, X., 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, 267–288. doi:10.3200/JEXE.74.3.267-288.
- SPSS. (2007). *SPSS 15.0 for Windows*. Chicago: SPSS, Inc.
- Steiger, J. H. (1989). *EzPath: A supplementary module for SYSTAT and SYGRAPH*. Evanston, IL: SYSTAT Inc.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180. doi:10.1207/s15327906mbr2502_4.
- Stern, M., Mazzeo, S. E., Gerke, C. K., Porter, J. S., Bean, M. K., & Laver, J. H. (2007). Gender, ethnicity, psychosocial factors, and quality of life among severely overweight, treatment-seeking adolescents. *Journal of Pediatric Psychology*, 32, 90–94. doi:10.1093/jpepsy/jsl013.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10. doi:10.1007/BF02291170.
- Uzark, K., Jones, K., Slusher, J., Limbers, C. A., Burwinkle, T. M., & Varni, J. W. (2008). Quality of life in children with heart disease as perceived by children and parents. *Pediatrics*, 121, e1060–e1067. doi:10.1542/peds.2006-3778.
- 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. doi:10.1177/109442810031002.
- Varni, J. W., Burwinkle, T. M., Berrin, S. J., Sherman, S. A., Artavia, K., Malcarne, V. L., et al. (2006a). The PedsQL™ in pediatric cerebral palsy: Reliability, validity, and sensitivity of the Generic Core Scales and Cerebral Palsy Module. *Developmental Medicine and Child Neurology*, 48, 442–449. doi:10.1017/S001216220600096X.
- Varni, J. W., Burwinkle, T. M., Jacobs, J. R., Gottschalk, M., Kaufman, F., & Jones, K. L. (2003a). The PedsQL™ in Type 1 and Type 2 diabetes: Reliability and validity of the Pediatric Quality of Life Inventory™ Generic Core Scales and Type 1 Diabetes Module. *Diabetes Care*, 26, 631–637. doi:10.2337/diacare.26.3.631.
- Varni, J. W., Burwinkle, T. M., Katz, E. R., Meeske, K., & Dickinson, P. (2002a). 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. doi:10.1002/cncr.10428.
- Varni, J. W., Burwinkle, T. M., Rapoff, M. A., Kamps, J. L., & Olson, N. (2004). The PedsQL™ in pediatric asthma: Reliability and validity of the Pediatric Quality of Life Inventory™ Generic Core Scales and Asthma Module. *Journal of Behavioral Medicine*, 27, 297–318. doi:10.1023/B:JOBM.0000028500.53608.2c.
- Varni, J. W., Burwinkle, T. M., & Seid, M. (2006b). The PedsQL™ 4.0 as a school population health measure: Feasibility, reliability, and validity. *Quality of Life Research*, 15, 203–215. doi:10.1007/s11136-005-1388-z.
- Varni, J. W., Burwinkle, T. M., Seid, M., & Skarr, D. (2003b). The PedsQL™ 4.0 as a pediatric population health measure: Feasibility, reliability, and validity. *Ambulatory Pediatrics*, 3, 329–341. doi:10.1367/1539-4409(2003)003<0329:TPAAPP>2.0.CO;2
- Varni, J. W., Limbers, C. A., & Burwinkle, T. M. (2007a). 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. doi:[10.1186/1477-7525-5-1](https://doi.org/10.1186/1477-7525-5-1).
- Varni, J. W., Limbers, C. A., & Burwinkle, T. M. (2007b). Parent proxy-report of their children's health-related quality of life: An analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL™ 4.0 Generic Core Scales. *Health and Quality of Life Outcomes*, 5(2), 1–10. doi:[10.1186/1477-7525-5-1](https://doi.org/10.1186/1477-7525-5-1).
- Varni, J. W., Seid, M., Knight, T. S., Burwinkle, T. M., Brown, J., & Szer, I. S. (2002b). The PedsQL™ in pediatric rheumatology: Reliability, validity, and responsiveness of the Pediatric Quality of Life Inventory™ Generic Core Scales and Rheumatology Module. *Arthritis and Rheumatism*, 46, 714–725. doi:[10.1002/art.10095](https://doi.org/10.1002/art.10095).
- 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. doi:[10.1097/00005650-200108000-00006](https://doi.org/10.1097/00005650-200108000-00006).
- 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. doi:[10.1097/00005650-199902000-00003](https://doi.org/10.1097/00005650-199902000-00003).
- von Rueden, U., Gosch, A., Rajmil, L., Bisegger, C., & Ravens-Sieberer, U. (2006). Socioeconomic determinants of health-related quality of life in childhood and adolescence: Results from a European study. *Journal of Epidemiology and Community Health*, 60, 130–135. doi:[10.1136/jech.2005.039792](https://doi.org/10.1136/jech.2005.039792).
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, 9, 1–26. doi:[10.1177/014662168500900101](https://doi.org/10.1177/014662168500900101).
- Willke, R. J., Burke, L. B., & Erickson, P. (2004). Measuring treatment impact: A review of patient-reported outcomes and other efficacy endpoints in approved product labels. *Controlled Clinical Trials*, 25, 535–552. doi:[10.1016/j.cct.2004.09.003](https://doi.org/10.1016/j.cct.2004.09.003).
- World Health Organization. (1948). *Constitution of the World Health Organization: Basic Document*. Geneva, Switzerland: World Health Organization.