

Agreement of Children and Parents Scores on Chinese Version of Pediatric Quality of Life Inventory Version 4.0: Further Psychometric Development

Chung-Ping Cheng¹ · Wei-Ming Luh² ·
Ai-Lun Yang³ · Chia-Ting Su⁴ · Chung-Ying Lin⁵

Received: 21 October 2014 / Accepted: 14 April 2015 / Published online: 24 April 2015

© Springer Science+Business Media Dordrecht and The International Society for Quality-of-Life Studies (ISQOLS) 2015

Abstract Quality of life (QoL) is an important index that allows health practitioners to understand the overall health status of an individual. One commonly used reliable and valid QoL instrument with parallel items on parent and child questionnaires, the Pediatric Quality of Life Inventory Version 4.0 (PedsQL), has been being developed since 1997. However, the use of parent- and child-reported PedsQL is still under development. Using multitrait-multimethod (MTMM) analyses and absolute agreement analyses across parent and child questionnaires can further help health practitioners understand the construct of PedsQL, and the feasibility of PedsQL in clinical. We analyzed the questionnaires of 254 parent–child dyads. MTMM through confirmatory factor analyses and percent of smallest real difference (SRD%) were used for analyzing. Our results supported the construct validity of the PedsQL. Four traits (physical, emotional, social, and school) and two methods (parent-proxy reports and child self-reports) were distinguished by MTMM. Moreover, the results of absolute agreements suggested that parent-rated and child-rated PedsQL are close (SRD% = 17.88–30.55 %); thus, a parent-rated PedsQL can be a secondary outcome representing a

✉ Chung-Ying Lin
cylin36933@gmail.com

¹ Department of Psychology, College of Social Science, National Cheng Kung University, Tainan, Taiwan

² Institute of Education, College of Social Science, National Cheng Kung University, Tainan, Taiwan

³ Graduate Institute of Exercise Science, University of Taipei, Taipei, Taiwan

⁴ Department of Occupational Therapy, College of Medicine, Fu Jen Catholic University, New Taipei City, Taiwan

⁵ Department of Public Health, College of Medicine, National Cheng Kung University, 1 University Road, Tainan 70101, Taiwan

child's health. We conclude that the PedsQL is useful for measuring children's QoL, and has helpful clinical implications.

Keywords Agreement · Confirmatory factory analysis · Multitrait-multimethod analysis · Smallest real difference · Quality of life

Introduction

Health practitioners use client-centered approaches to improve an individual's health (Lin et al. 2012b), and they need effective evaluations to assess the health condition. Thus, quality of life (QoL) measurement tools, which provide useful and subjective information (World Health Organization [WHO] 1993; Young et al. 2013), can yield an effective evaluation for healthcare practitioners. In addition, QoL measures have been developed for various populations, including the children (Ravens-Sieberer et al. 2006; Roy et al. 2013). Children, however, may lack the cognitive skills necessary to accurately complete a QoL questionnaire; therefore, parent-proxy reports have been developed for some pediatric QoL instruments (Ravens-Sieberer and Bullinger 1998, 2000; Varni et al. 1999, 2001, 2008).

Some evidence has shown that even children younger than 8 years can use rating scales, understand response terms and the underlying concepts, and assess their own QoL (Cremeens et al. 2006a, b). Therefore, using child self-reports is better than using parent-proxy reports to measure a child's QoL. However, even though a self-report is better than a proxy-report, a parent-proxy report provides valuable supplemental information for health practitioners. Firstly, concurrent parent-proxy reports and child self-reports can be used to examine the construct validity of QoL instruments. Because QoL is multidimensional (WHO 1993), items on QoL questionnaires are supposed to be discriminant for different dimensions, and to be convergent for the same dimension. The construct validity (i.e., discriminant and convergent validity) for the different and same dimensions can be examined by using two measuring methods (e.g., parent-proxy reports and child-self reports). The test for measuring the validity of these methods is called multitrait-multimethod (MTMM) analysis. Correlations of the same trait (the same dimension in QoL) between different methods are supposed to be higher than those of different traits (different dimensions in QoL) between different methods (Campbell and Fiske 1959). Furthermore, MTMM analysis that uses confirmatory factor analysis (CFA) models can provide thorough evidence of construct validity (Huang and Michael 2002; Marsh and Grayson 1995) and a quantitative description of the method effect for QoL instruments. Because the instrument has been validated, healthcare practitioners can evaluate the patients' health conditions and assess the effectiveness of their interventions.

The other advantage of parent-proxy reports is that healthcare practitioners can know the parents' perceptions of their children's QoL, and can compare the agreement between parents and children (Eiser and Morse 2001; Lin et al. 2013b; Upton et al. 2008). After knowing the agreement, healthcare practitioners may help to improve the children's QoL (Su et al. 2013). For example, when parents overrate their children's QoL, healthcare practitioners can help parents understand their children's QoL difficulties. Likewise, when parents underrate their children's QoL, healthcare practitioners

can help parents understand that their children are doing well and that they need not pressure their children to improve their QoL. In addition, parent-proxy reports can be the primary outcome measure when the child is too young or too ill to complete a self-report (Limbers et al. 2008), especially if agreement across parent-proxy reports and child self-reports is established. Therefore, knowing the relationships between parent- and child-reported questionnaires would be helpful for healthcare practitioners making clinical decisions.

The primary methods for comparing agreement between parents and children are correlation analyses, such as the Pearson correlation, Spearman correlation, and intraclass correlation (ICC). However, these correlation analyses measure only relative agreement; a high correlation reveals little information regarding agreement (Chen et al. 2007). In contrast, absolute agreement indices, including standard error of measurement (SEM) and smallest real difference (SRD), help healthcare practitioners understand the variances that are not caused by traits (Chen et al. 2007; Su et al. 2014). In other words, SEM and SRD provide the information of variances from different methods or from measurement errors. Moreover, SEM and SRD represent the agreement for a group of individuals and for a single individual, respectively. Therefore, SEM indicates how the ratings of a group of parents agree with their children's ratings on QoL, and SRD indicates how the QoL rating of a parent agrees with that of his or her child. A lower SEM and SRD indicate better agreement. In addition to SEM and SRD, the Bland-Altman method provides a visualization of parent-child agreement (Rankin and Stokes 1998), and is helpful for healthcare practitioners to understand the relationship between child- and parent-rated reports.

To the best of our knowledge, no studies have tested the psychometric properties of a commonly used generic QoL questionnaire (Pediatric Quality of Life Inventory Version 4.0 [PedsQL]) with MTMM analyses using CFA. In addition, absolute agreements have not been examined for PedsQL. Because an instrument should be evaluated in several ways to establish its validity, this study focused on two purposes that correspond to psychometric developing. One purpose was to use two types of MTMM analysis: the traditional one and the one with CFA, to investigate the construct validity of the PedsQL. Another purpose was to assess both the relative and the absolute agreement across parent-proxy reports and child self-reports.

Methods

Sample

We used secondary data, for which the sampling and procedures have been reported elsewhere (Lin et al. 2013a). In brief, information about this study was disseminated by teachers of 11 elementary schools, and the children and parents interested in participating in this study signed informed consents. Those children who agreed to participate filled out the child self-report in their classrooms under the supervision of their teachers and one researcher. The parent-proxy reports were taken back home by the children for one of their parents to

complete, and the parent reports were collected from the children 1 to 3 days later. In the current study, we included only 254 parent–child dyad questionnaires that had no missing values.

Quality of Life Questionnaire: Pediatric Quality of Life Inventory Version 4.0 (PedsQL)

The Chinese version of the short form PedsQL is a validated 15-item questionnaire for children aged 8 to 12 years (Lin et al. 2012a, 2013a). In addition, evidence supports the notion that comparisons between parents' and children's reports are meaningful, that is, that measurement equivalence has been established across parent reports and child reports of PedsQL (Lin et al. 2013a). Each item on the PedsQL asks the frequency of a difficulty's occurrence. These frequencies are then transformed into a 0–100 scale (never = 100, almost never = 75, sometimes = 50, often = 25, and almost always = 0). Four subscale scores, two summary scores, and one total score are computed in parent- and child-reported PedsQL separately and according to the developer's instructions (Chan et al. 2005; Varni et al. 2001). The physical summary score is the same as the physical subscale, and the psychosocial summary score is the average of the other three subscales. Higher scores represent better QoL for the corresponding subscales, summary scores, and total score.

Data Analysis

For MTMM analyses, Pearson correlation coefficient, Cronbach's α , and CFA models were used. An 8×8 Pearson correlation matrix was created, and the correlation effects were designated as small (.1 to .29), medium (.3 to .49), and large ($\geq .5$) (Varni et al. 2001). A Cronbach's α value $>.7$ is regarded as acceptable (Portney and Watkins 2000). In addition, five competing CFA models were used to explore the structure of the MTMM matrix and the effects of traits and methods on items. Model 1 is a four-oblique-trait-factor (physical, emotional, social, and school factors) model (Fig. 1a); Model 2¹ is a two-oblique-method-factor (parent- and child-reported PedsQL) model (Fig. 1b); and Model 3 is a one-general-factor (the overall QoL factor) and two-oblique-method-factor model (Fig. 1c). Model 4 is a four-oblique-trait-factor and two-method-factor model (Fig. 1d). Model 5 is a four-oblique-trait-factor and two-minus-one-method-factor model (Fig. 1e). Then, we compared Models 1 with 4 for testing the method effects, compared Models 2 with 3 and Models 2 with 4 for examining the convergent validity, and compared Models 3 with 4 for understanding the discriminant validity between hypothetical traits given method effect is considered. Model 5 was used to compensate one major limitation of our Model 4. That is, slightly insufficient *degrees of freedom* ($df = 5$ in our

¹ Please note that Model 2 also provides the information of agreement between parents and children. However, we used other statistical methods (including ICC, SEM, and SRD) to examine the agreement based on the reason that Model 2 can only provide the overall agreement. In other words, Model 2 provides the agreement of the entire QoL, while other statistical methods we used provide the agreement for each QoL domain (e.g., physical and psychosocial domains).

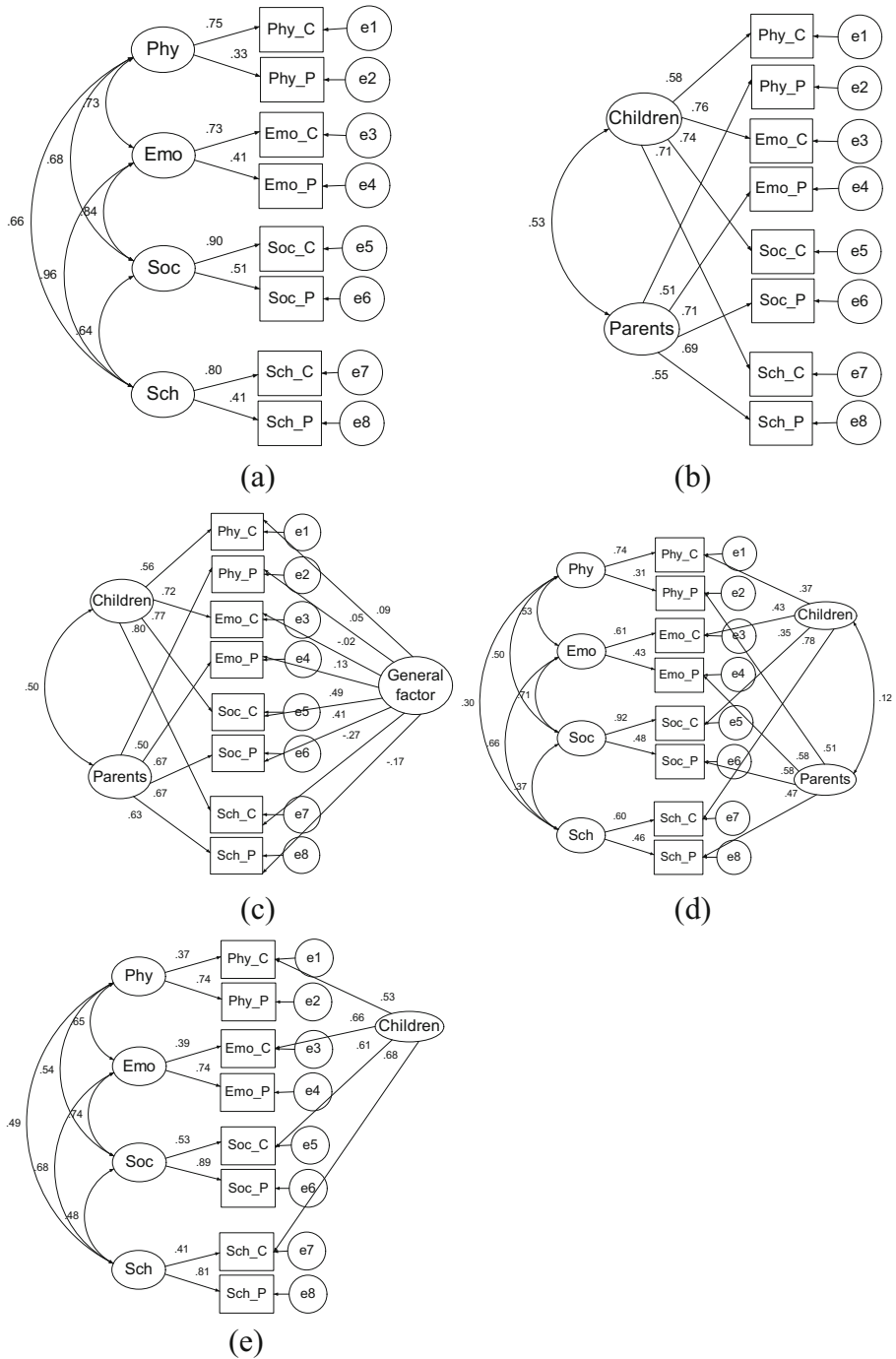


Fig. 1 *Phy* = physical; *Emo* = emotional; *Soc* = social; *Sch* = school; *P* = parent; *C* = children. **Model 1:** Four oblique trait-factors model with standardized coefficients. **Model 2:** Two oblique method-factors model with standardized coefficients. **Model 3:** One general factor and two oblique method-factors with standardized coefficients. **Model 4:** Four oblique trait-factors and two oblique method-factors model with standardized coefficients

Model 4) that might lead to convergence problem. On the other hand, although Model 5 only contains one method effect (the child reports in our case), it also tested both trait and method effects as Mode 4 did. Because all the scores in subscales can be seen as normally distributed (all absolute skewnesses < 3 and all kurtoses < 10; Table 1) (Kline 1998) in both parent- and child-reported PedsQL, using the maximum likelihood estimation for CFA was appropriate.

In these competing CFA models, nine goodness-of-fit indices were used to evaluate how the data fit the hypothesized model. For the models, a χ^2 test that shows no significant difference between the hypothesized model and the data suggests a satisfactory fit. However, a χ^2 test is easily biased when the sample size is large; therefore, other indices are recommended instead (Hoyle and Panter 1995; Schweizer 2010). Consequently, we used χ^2/df , the standardized root mean square residual (SRMR), the root mean square error of approximation (RMSEA), the goodness of fit index (GFI), the adjusted goodness of fit index (AGFI), the comparative fit index (CFI), the

Table 1 Descriptive statistics of parent- and child-reported PedsQL ($N=254$)

	No. of items	Mean	SD	Skewness	Kurtosis	Comparisons between parent report and child report				
						Difference	ICC	SEM	SRD	SRD%
Parent report										
Physical	5	90.06	11.41	-2.13	6.03	1.13	.398	7.52	20.83	20.83
Emotional	4	83.58	16.10	-0.97	0.60	2.95	.452	11.02	30.55	30.55
Social	3	90.59	14.30	-2.21	7.09	1.50	.621	8.38	23.23	23.23
School	3	81.08	14.59	-0.97	1.02	-2.81	.487	8.86	24.56	24.56
Psychosocial ^a	10	85.09	11.74	-0.89	0.74	0.54	.580	7.32	20.29	20.29
Total ^b	15	86.81	10.31	-0.93	0.61	0.75	.555	6.45	17.88	17.88
Child report										
Physical	5	88.93	12.01	-1.84	3.53					
Emotional	4	80.63	20.50	-1.32	1.68					
Social	3	89.09	17.18	-2.14	4.99					
School	3	83.90	14.80	-1.63	3.70					
Psychosocial ^a	10	84.54	14.81	-1.65	3.69					
Total ^b	15	86.07	12.62	-1.59	3.31					

Difference Parent-reported scores–child-reported scores

SEM standard error of measurement = standard deviation of the parent–child mean score $\times \sqrt{(1-ICC)}$

SRD smallest real difference = $1.96 \times SEM \times \sqrt{2}$

SRD% SRD/100, and the values <30 % are in **bold font**

ICC Intraclass correlation coefficient; values >.4 are in **bold font**

* $p < .05$

^a Psychosocial = average of emotional, social, and school

^b Total = average of all 15 items

incremental fit index (IFI), and the Tucker-Lewis index (TLI) to examine the model fit. A $\chi^2/df < 3$, an SRMR and a RMSEA $< .08$, and other indices $> .9$ suggested a satisfactory data-model fit (Hoyle and Panter 1995; Schweizer 2010).

Relative agreement was examined using ICC²: $>.75$ was considered excellent; between $.75$ and $.4$, fair-to-good; and $< .4$, poor (Rosner 2006). Absolute agreement was examined using SEM³ (= standard deviation of the parent-child mean score $\times \sqrt{[1 - \text{ICC}]}$) and SRD⁴ ($= 1.96 \times \text{SEM} \times \sqrt{2}$). In addition, SRD%⁵ (= SRD/total score range: SRD/100 in this study) was also calculated: an SRD% $< 30\%$ was considered good agreement (Chen et al. 2009).

Descriptive data, Cronbach's α , MTMM using Pearson's correlation, and ICC were analyzed using SPSS 16.0 (SPSS Inc., Chicago, IL, USA). MTMM combined with CFA models were analyzed using AMOS 7.0 (SPSS Inc.). SEM, SRD, and SRD% were computed using Microsoft Excel 2007.

Results

Participant Characteristics

The 254 children were between 9 and 12 years old (mean \pm SD: 10.68 \pm 1.18 years), and 120 of them were girls (47.2 %). For their parents, mothers ($n=200$) and fathers ($n=$

² The ICC shares the same terms of variances in the repeated measures analysis of variance (ANOVA): variance between participants and that within participants. Of the variance within participants, it can be separated as variance between methods and residual variance. Therefore, we could have the following model: $x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, where β represents for participants, α for methods (i.e., parent- and child-rated PedsQL in this study), and ε for residual. Then, the ICC can be computed as β variance divided by the sum of α , β , and ε variances. Based on the formula, we could easily know that a high value of ICC shows a fair degree of agreement between the methods. In addition, ICC simultaneously accounts for bias (i.e., whether children rate the PedsQL lower or higher than parents do) and association (i.e., whether children and parents understand the meaning of the PedsQL in the same way).

³ SEM equals to the square root of the error variance: $\text{SEM} = \sqrt{\sigma^2_{\text{error}}} = \sqrt{\sigma^2_{\text{methods}} + \sigma^2_{\text{residual}}}$, where $\sigma^2_{\text{methods}}$ represents for the variance of child- and parent-rated PedsQL. Because ICC is computed as $\sigma^2_{\text{methods}}$ divided by total σ^2 , $1 - \text{ICC} = \sigma^2_{\text{residual}}$ divided by total σ^2 . SEM then can be calculated as: $\text{SEM} = \sigma_{\text{total}} \times \sqrt{(1 - \text{ICC})}$. Hence, based on the formula of SEM, we could examine the differences between child- and parent-rated PedsQL.

⁴ $\text{SRD} = 1.96 \times \text{SEM} \times \sqrt{2}$, where 1.96 represents the 95 % confidence interval from a normal distribution, and $\sqrt{2}$ is used to account for the additional uncertainty introduced by using difference scores from the 2 independent measurements with the same variances (in our study, they are child-rated and parent-rated PedsQL). The variance of the difference scores ($\text{SD}^2_{\text{diff}}$) can be computed from three sources: variances of the child-rated ($\text{SD}^2_{\text{child}}$) and parent-rated ($\text{SD}^2_{\text{parent}}$) scores, and the covariance ($\text{cov}_{\text{child, parent}}$) between them. Therefore, $\text{SD}^2_{\text{diff}} = \text{SD}^2_{\text{child}} - 2 \times \text{cov}_{\text{child, parent}} + \text{SD}^2_{\text{parent}}$. Because the Pearson correlation (r) can be defined as the ratio of the covariance divided by the product of the corresponding SDs (i.e., $\text{SD}_{\text{child}} \times \text{SD}_{\text{parent}}$), $\text{cov}_{\text{child, parent}}$ is substituted by $r \times \text{SD}_{\text{child}} \times \text{SD}_{\text{parent}}$. And $\text{SD}^2_{\text{diff}} = \text{SD}^2_{\text{child}} - 2 \times r \times \text{SD}_{\text{child}} \times \text{SD}_{\text{parent}} + \text{SD}^2_{\text{parent}}$. Assuming child-rated and parent-rated PedsQL have equal variability in the population, we can get $\text{SD}^2_{\text{diff}} = 2 \times \text{SD}^2_{\text{child}} - 2 \times r \times \text{SD}^2_{\text{child}} = 2 \times \text{SD}^2_{\text{child}} \times (1 - r)$. Taken square root in both sides: $\text{SD}_{\text{diff}} = \text{SD}_{\text{child}} \times \sqrt{2} (1 - r)$. Assuming child-rated and parent-rated PedsQL are independent, the Pearson correlation r is zero, and the last term is reduced to $\text{SD}_{\text{diff}} = \text{SD}_{\text{child}} \times \sqrt{2}$. Hence, using $\sqrt{2}$ to multiply SEM is to conservatively consider the possibly largest uncertainty.

⁵ The primary benefit of using SRD% is that it is independent of the units of measurement, and readers can easily understand the magnitude of a bias is. Another benefit of using the total range is that total range is the same across samples, while standard deviation will be changed in different samples.

196) were 39.95 ± 5.29 and 42.69 ± 6.45 years, respectively. The proxy raters were mothers ($n=159$), fathers ($n=82$), others (e.g., grandparents) ($n=10$), and anonymous ($n=3$). The QoL scores are presented on Table 1.

Multi-Traits Multi-Methods of PedsQL

The correlation coefficients ranged from small-to-medium for monotrait-heteromethod (same QoL dimension in different methods, $r=.25$ to $.46$) and heterotrait-heteromethod values (different QoL dimensions in different methods, $r=.11$ to $.32$), and were from medium-to-large for heterotrait-monomethod values (different QoL dimensions in the same method, $r=.31$ to $.58$). To examine discriminant validity, Campbell and Fiske (1959) suggest comparing monotrait-heteromethod values both with heterotrait-heteromethod values and with heterotrait-monomethod values. Among 24 comparisons, 23 (95.8 %) monotrait-heteromethod values were higher than heterotrait-heteromethod values, but only 3 (12.5 %) values were higher than heterotrait-monomethod values. The Cronbach’s α values were all higher than .7 in both parent- (Physical: .77, Emotional: .84, Social: .81, and School: .80) and child-reported PedsQL (Physical: .73, Emotional: .83, Social: .83, and School: .74) (Table 2).

The best data-model fit for the five competing models, i.e., the models in which all indices met the criteria, were Models 4 and 5 (Table 3). The other three models did not meet the criteria in at least one case (Model 3) and in as many as nine (Model 1). Figure 1 shows the standardized factor loadings and correlation of Models 1 to 4.

Table 2 Multi-traits multi-methods correlation for PedsQL ($N=254$)

	Parent report				Child report			
	Phy	Emo	Soc	Sch	Phy	Emo	Soc	Sch
Parent report								
Phy	<i>(.77)</i>							
Emo	.37**	<i>(.84)</i>						
Soc	.35**	.49**	<i>(.81)</i>					
Sch	.31**	.38**	.36**	<i>(.80)</i>				
Child report								
Phy	<u>.25**</u>	<i>.18**</i>	<i>.21**</i>	<i>.12*</i>	<i>(.73)</i>			
Emo	<i>.15*</i>	<u>.30**</u>	<i>.23**</i>	<i>.23**</i>	.40**	<i>(.83)</i>		
Soc	<i>.16*</i>	.32**	<u>.46**</u>	<i>.17**</i>	.47**	.54**	<i>(.83)</i>	
Sch	<i>.11</i>	.22**	<i>.16*</i>	<u>.32**</u>	.42**	.58**	.48**	<i>(.74)</i>

Cronbach’s α values are in parentheses, heterotrait-heteromethod values are in *italic font*, heterotrait-monomethod values are in **bold font**, monotrait-heteromethod values are underlined

Phy Physical; Emo Emotional; Soc Social; Sch School

* $p < .05$; ** $p < .01$

Table 3 Goodness of fit indices for competing models ($N=254$)

	Model #					Model comparison				
	1	2	3	4	5	1-4	2-4	3-4	5-4	
χ^2	149.40*	75.51*	20.00*	2.19	10.64	$\Delta\chi^2$	147.21*	73.31*	17.81*	8.45
df	14	19	11	5	10	Δdf	9	14	6	5
χ^2/df	10.671	3.974	1.818	.438	1.064					
SRMR	.109	.053	.029	.010	.023	Δ SRMR	.099	.043	.019	.013
RMSEA	.196	.108	.057	.000	.016	Δ RMSEA	.196	.108	.057	.016
GFI	.845	.931	.981	.998	.990	Δ GFI	.153	.067	.017	.008
AGFI	.600	.870	.938	.985	.963	Δ AGFI	.385	.115	.047	.022
CFI	.756	.898	.984	1.000	.999	Δ CFI	.244	.102	.016	.001
IFI	.762	.900	.984	1.005	.999	Δ IFI	.243	.105	.021	.006
TLI	.511	.850	.959	1.028	.997	Δ TLI	.517	.178	.069	.031

Model 1: Four-oblique-trait-factor model

Model 2: Two-oblique-method-factor model

Model 3: One-general-factor and two-oblique-method-factor model

Model 4: Four-oblique-trait-factor and two-oblique-method-factor model

Model 5: Four-oblique-trait-factor and two-minus-one-method-factor model

df degrees of freedom; SRMR Standardized root mean square residual; RMSEA Root mean square error of approximation; GFI Goodness of fit index; AGFI Adjusted goodness of fit index; CFI Comparative fit index; IFI Incremental fit index; TLI Tucker-Lewis index

* $p < .05$

We then compared Model 4 with the other four.⁶ The method effects, convergent validity, and discriminant validity were supported by the significant $\Delta\chi^2$ of Model 4 versus 1, versus 2, and versus 3 (Table 3). In addition, all the fit indices in Model 4 were better than those in the other three. The child reports had much higher standardized coefficients for each factor than did the parent reports. Moreover, no significant difference was found between Models 4 and 5, and suggests the two models are similar.

Agreement Between Parents and Children

The relative agreements of all the subscale scores (ICC = .452 to .621) and the total score (ICC = .555) between parents and children were from fair-to-good, except for the physical subscale, which was only close to fair (ICC = .398) (Table 1). In addition, the absolute agreements showed that all the subscale

⁶ Model 1 only accounts for QoL trait; Model 2 only accounts for different methods (i.e., child- and parent-rated PedsQL); Model 3 accounts for one general QoL trait and two methods; Models 4 and 5 account for both four QoL traits and two methods. Therefore, comparing Models 4 and 1 helps us understand the method effects; Models 4 and 2 helps us examine the trait effects, that is, whether child- and parent-rated QoL were satisfactory converged, and indicates convergent validity. The difference between Models 4 and 3 is that Model 4 used QoL as separated traits, while Model 3 used QoL as one general trait. Therefore, comparing the two models help we understand the discriminant between the four QoL traits, and discriminant validity can be tested.

scores (SRD% = 20.29–24.56) and the total score (SRD% = 17.88) were satisfactory, except for the emotional subscale, which was only close to acceptable (SRD% = 30.55). Moreover, the Bland-Altman figures showed rightward arrows in all the subscales and the total score, and most parent–child dyads fell within the 95 % limits of agreement (mean difference \pm 1.96 SD: the horizontal lines immediately above and below the center line at 0 00) (Fig. 2).

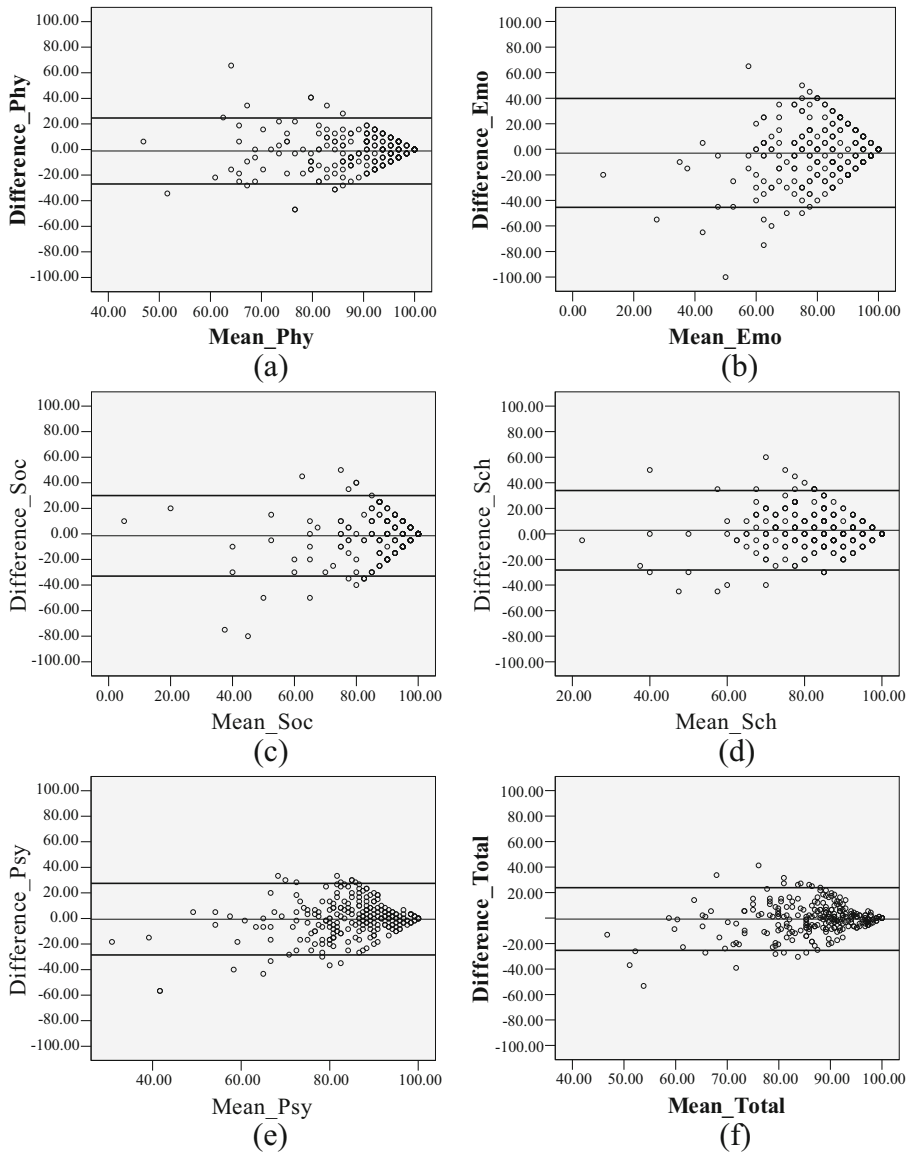


Fig. 2 The Bland-Altman figures of parent- and child-rated PedsQL. *Phy* = physical; *Emo* = emotional; *Soc* = social; *Sch* = school; *Psy* = psychosocial; *Total* = total score; *Mean* = (parent-rated PedsQL score + child-rated PedsQL score)/2; *Difference* = parent-rated PedsQL score – child-rated PedsQL score

Discussion

Our findings support the construct validity of the Chinese version of the PedsQL and of the relationship between parent- and child-rated reports in a Taiwan sample. Method effects, discriminant validity, and convergent validity were evidenced by our competing CFA models. In addition, relative and absolute agreements between parents and children were acceptable in all of the QoL domain scores and in the total score.

Method Effects and the Construct Validity of the PedsQL

The MTMM matrix in this study is comparable to that in Varni et al. (2001), who reported medium correlation coefficients for the monotrait-heteromethod ($r=.36$ to $.50$) and heterotrait-monomethod values ($r=.42$ to $.49$), and small-to-medium correlation coefficients for heterotrait-heteromethod values ($r=.17$ to $.33$). All of their monotrait-heteromethod values were higher than all of their heterotrait-heteromethod values in 24 comparisons (100 %; our result: 95.8 %), and six of their monotrait-heteromethod values were higher than heterotrait-monomethod values in 24 comparisons (25 %; our result: 12.5 %). Although their results seem better than ours, both sets of results demonstrated the method effects of the PedsQL as evidenced by medium correlation in heterotrait-monomethod values and few monotrait-heteromethod values higher than heterotrait-monomethod values. Method factors rather than trait factors make observed items more correlated. We hypothesize that this method effect is caused by the halo effect (Thorndike 1920) in parents' reports. In other words, when parents believe that their children have a good physical QoL, they also tend to think that their children have a good emotional, social, and school QoL. Thus, the halo effect raises the heterotrait-monomethod values above those of the monotrait-heteromethod values.

The CFA models in the present study demonstrated the method effects, and supported that the discriminant and convergent validity of the PedsQL is valid and reliable for measuring children's QoL in various language versions (e.g., Chan et al. 2005; Huguet and Miró 2008; Kobayashi and Kamibeppu 2010; Upton et al. 2005), including the Chinese version used in Taiwan (Lin et al. 2012a, 2013a). However, although both parent-proxy reports and child-self reports are used, most of these studies examine the validity of PedsQL separately for these two methods. To the best of our knowledge, only Varni et al. (2001) has examined the construct validity by using the two methods at the same time, i.e., by using MTMM analysis, which is known to have several disadvantages when using the criteria to judge discriminant and convergent validity (Marsh and Grayson 1995). Campbell and Fiske (1959) said that the MTMM criteria should be viewed as "common-sense desideratum", and Marsh and Grayson (1995), that the criteria were not a "summative evaluation" or "global summaries" of method effects, discriminant, and convergent validity. Therefore, the alternative approach to MTMM matrices, CFA models, is widely used because they provide a thorough examination of method effects and of discriminant and convergent validity (Marsh and Grayson 1995; Tomás et al. 2000). The present study extends the evidence of method effects, discriminant validity, and convergent validity for the PedsQL by using the CFA approach to MTMM methodology.

In addition, we found that the standardized coefficients of child-self report scores are higher than those of parent-proxy report scores. Items in a child-self report are better

indicators of the traits we are interested in than are those in a parent-proxy report. Our findings support the notion of using child-self reports as the primary outcome for measuring children's QoL (Limbers et al. 2008). We could also see that the two methods were correlated mainly due to they share the same QoL traits. When the traits were not included in our model (see Model 2; Fig. 1b), both methods are moderately correlated ($r=.53$). However, when the traits and methods were simultaneously taken into account, i.e., traits and methods were separately extracted (see Model 4; Fig. 1d), both methods are low correlated ($r=.12$). Therefore, the correlations between child- and parent-rated PedsQL were mainly contributed from their same QoL traits. Moreover, the best CFA model, that is, Model 4, showed unusually satisfactory fit indices (e.g., CFI = 1; IFI and TLI over 1; RMSEA = 0), and some may speculate the results of Model 4. However, our Model 5 as a compensate model for Model 4 showed acceptable fit indices (All CFI, IFI, and TLI <1, and RMSEA>0); hence, we are confident to our demonstrated method effects, discriminant and convergent validity in the PedsQL.

Agreements Across Parents and Children

Upton et al. (2008) pointed out that of 10 studies using PedsQL as the measurement, five found higher parent-child agreement for the physical part (i.e., concrete, observable characteristics), four others found higher levels for psychosocial domains, and one (Upton et al. 2005) found better agreement on the physical subscale only when the children were unhealthy and reported the correlation as .2 in the physical domain, and as .27 to .42 in the psychosocial domains for healthy children. Because our participants had no chronic illnesses (they were all healthy), our findings of better relative agreements in the psychosocial domain scores (ICC = .452 to .621, $r=.30$ to .46) than those in the physical domain score (ICC = .398, $r=.25$) agree with the findings of Upton et al. (2005). Cremmens et al. (2006a, b) also share a similar trend in a healthy sample: ICC = .12 in psychosocial domain, and = .02 in physical domain. While the relatively low ICC values in Cremmens et al's study as compared with ours may be due to their young participants (age 5.5 to 8.5 years). Using a stable physically disabled sample, Sheffler et al. (2009) also found higher ICC values in psychosocial domain (.36 to .47) as compared with that in physical domain (.34). The ICC values in psychosocial domain (.31 to .74) also seemed to be higher than that in physical domain (.34) of an overweight sample (Lin et al. 2013b). In addition, our finding of an acceptable SRD% was comparable to that of Upton et al. (2008), who found only small and insignificant differences between parent-proxy report scores and child-self report scores. Moreover, the Bland-Altman figures show that parent-child agreements were higher in children with a better than a worse QoL.

Clinical Implications

Our findings suggest several implications for health practitioners. First, it is feasible to use the PedsQL to rate children's QoL and to evaluate their QoL difficulties. Because we found that the factor structures of the PedsQL's parent-proxy reports and child self-reports are consistent, we suggest using both at the same time if both can be applied. In addition, parents and children should not fill out their questionnaires together because it

may increase their agreement (Upton et al. 2008). We recommend taking the child-rated QoL as the primary QoL outcome, and the parent-rated as the secondary outcome. Second, comparing the difference between parent- and child-rated QoL can also provide health practitioners with information that they can use to understand whether parents know their children's QoL difficulties, and that should suggest whether they ought to intervene when, for example, a large difference is found. Third, when the child is unable to fill out the PedsQL, health practitioners can use the parent-proxy report to evaluate the child's QoL, but the discrepancy between the two forms should be taken into account. Although the parent-proxy report might provide the health practitioners some insight of the child's QoL, the parent-rated QoL cannot completely substitute the child-rated QoL. For example, our results showed that parents tend to overestimate their children's QoL in Physical, Emotional, and Social domains; while underestimated in School domain. The overestimation and underestimation should be concerned when use a parent-proxy report to estimate a child's QoL. Moreover, if a parent rates a low QoL for his or her child, health practitioners could be alerted that the child may have some difficulties. However, this suggests that health practitioners should additionally and thoroughly examine the child instead of treating the child mainly based on the parent-reported QoL.

Limitations and Future Directions

This study had some limitations. First, convenience sampling was used to recruit the participants (Lin et al. 2013a), which may limit the generalizability of our results. Second, the parents who participated in this study may pay more attention to their children than those who did not participate; thus, the results of discrepancies and agreements in this study may be limited. Because all the parents knew that the PedsQL was an instrument that would help them understand their children's health, we assume that the parents who completed the questionnaire may care more about their children than those who did not. However, it is impossible for us to know the ratings from parents who do not participate. Moreover, when we compared the child-reported PedsQL scores of children whose parents filled out parent-proxy reports with those of children whose parents did not, we found no significant differences. Therefore, parents who did not complete the parent-proxy report likely had little or no effect on our results. Third, different kinds of parent-proxy ratings (from fathers, mothers, grandparents, and other caregivers) were used in this study. Jozefiak et al. (2008) reported only moderate agreement between mother-rated and father-rated proxy reports. In addition, they pointed out that their findings may be applicable only for countries with a gender-role structure similar to those in Sweden and Norway. Therefore, further studies are needed on the differences between the proxy ratings given by different types of raters of children's QoL. Fourth, we only used two methods (i.e., parent- and child-reported PedsQL) to perform MTMM, while three or more methods are recommended to strongly support the convergent and discriminant validity (Campbell and Fiske 1959; Kenny and Kashy 1992; Marsh and Bailey 1991). Therefore, a third method is suggested for future studies to examine the validity of the PedsQL. For example, a teacher-reported PedsQL might be useful. Fifth, although the PedsQL can be used to assess children with chronic illness, this study did not recruit any children with a chronic illness. The discrepancy and agreement across parent-proxy reports and child

self-reports may be different between children with and without a chronic illness. Therefore, future studies using the PedsQL to understand the discrepancy and agreement for children with a chronic illness are suggested. Moreover, because our study only recruited participants in Taiwan, the generalizability of our results may confine to the following East Asia countries/cities, Japan, South Korea, People's Republic of China, and Hong Kong, which share similar culture or traditional value with our sample. All parents of the countries/cities highly respect academic achievement, and most of the school-age children in the countries/cities go to cram schools after public or private school (Lin et al. 2012b). Moreover, the mentioned countries/cities are collectivism and emphasize “we” mentality and group solidarity in societies (Chun et al. 2005; Hofstede 2001; Trandis 1995). However, future studies are still needed to corroborate our results.

Acknowledgments This research was, in part, supported by the Ministry of Education, Taiwan, R. O. C. The Aim for the Top University Project to the National Cheng Kung University (NCKU).

References

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.
- Chan, K. S., Mangione-Smith, R., Burwinkle, T. M., Rosen, M., & Varni, J. W. (2005). The PedsQL: reliability and validity of the short-form generic core scales and asthma module. *Medical Care*, *43*(3), 256–265.
- Chen, H.-M., Hsieh, C.-L., Lo, S. K., Liaw, L.-J., Chen, S.-M., & Lin, J.-H. (2007). The test-retest reliability of 2 mobility performance tests in patients with chronic stroke. *Neurorehabilitation and Neural Repair*, *21*(4), 347–352.
- Chen, H.-M., Chen, C.-C., Hsueh, I.-P., Huang, S.-L., & Hsieh, C.-L. (2009). Test-retest reproducibility and smallest real difference of 5 hand function tests in patients with stroke. *Neurorehabilitation and Neural Repair*, *23*(5), 435–440.
- Chun, C.-A., Moos, R. H., & Cronkite, R. C. (2005). Culture: a fundamental context for the stress and coping paradigm. In P. T. P. Wang & L. C. J. Wong (Eds.), *Handbook of multicultural perspectives on stress and coping* (pp. 29–53). New York: Springer.
- Cremeens, J., Eiser, C., & Blades, M. (2006a). Characteristics of health-related self-report measures for children aged three to eight years: a review of the literature. *Quality of Life Research*, *15*(4), 739–754.
- Cremeens, J., Eiser, C., & Blades, M. (2006b). Factors influencing agreement between child self-report and parent proxy-reports on the Pediatric Quality of Life Inventory™ 4.0 (PedsQL™) generic core scales. *Health and Quality of Life Outcomes*, *4*, 58.
- Eiser, C., & Morse, R. (2001). Can parents rate their child's health-related quality of life? Results of a systematic review. *Quality of Life Research*, *10*(4), 347–357.
- Hofstede, G. (2001). *Culture's consequences* (2nd ed.). Thousand Oaks: Sage Publications.
- Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and application*. Thousand Oaks: Sage.
- Huang, C.-J., & Michael, W. B. (2002). Multitrait-multimethod analyses of scores on a Chinese version of the dimensions of self-concept scale. *Educational and Psychological Measurement*, *62*(2), 355–372.
- Huguet, A., & Miró, 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*(1), 63–79.
- Jozefiak, T., Larsson, B., Wichstrom, L., Mattejat, F., & Ravens-Sieberer, U. (2008). Quality of life as reported by school children and their parents: a cross-sectional survey. *Health and Quality of Life Outcomes*, *6*, 34.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112*, 165–172.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

- Kobayashi, K., & Kamibeppu, K. (2010). Measuring quality of life in Japanese children: development of the Japanese version of the PedsQL. *Pediatrics International*, *52*, 80–88.
- 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*(4), 659–668.
- Lin, C.-Y., Luh, W.-M., Yang, A.-L., Su, C.-T., Wang, J.-D., & Ma, H.-I. (2012a). Psychometric properties and gender invariance of the Chinese version of the self-report Pediatric Quality of Life Inventory Version 4.0: short form is acceptable. *Quality of Life Research*, *21*(1), 177–182.
- Lin, C.-Y., Su, C.-T., & Ma, H.-I. (2012b). Physical activity patterns and quality of life of overweight boys: a preliminary study. *Hong Kong Journal of Occupational Therapy*, *22*(1), 31–37.
- Lin, C.-Y., Luh, W.-M., Cheng, C.-P., Yang, A.-L., Su, C.-T., & Ma, H.-I. (2013a). Measurement equivalence across child self-reports and parent-proxy reports in the Chinese version of the Pediatric Quality of Life Inventory Version 4.0. *Child Psychiatry and Human Development*, *44*(5), 583–590.
- Lin, C.-Y., Su, C.-T., Wang, J.-D., & Ma, H.-I. (2013b). Self-rated and parent-rated quality of life for community-based obese and overweight children. *Acta Paediatrica*, *102*(3), e114–e119.
- Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analysis of multitrait-multimethod data: a comparison of alternative models. *Applied Psychological Measurement*, *15*, 47–70.
- Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Thousand Oaks: Sage.
- Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed., p. 708). Upper Saddle River: Prentice Hall Health.
- Rankin, G., & Stokes, M. (1998). Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clinical Rehabilitation*, *12*(3), 187–199.
- Ravens-Sieberer, U., & Bullinger, M. (1998). Assessing health-related quality of life in chronically ill children with the German KINDL: first psychometric and content analytical results. *Quality of Life Research*, *7*(5), 399–407.
- Ravens-Sieberer, U., & Bullinger, M. (2000). *KINDL^R-Questionnaire for measuring health-related quality of life in children and adolescents revised version manual*. <http://www.kindl.org/english/manual/>. Accessed 17 April 2015.
- Ravens-Sieberer, U., Erhart, M., Wille, N., Wetzel, R., Nickel, J., & Bullinger, M. (2006). Generic health-related quality-of-life assessment in children and adolescents: methodological considerations. *PharmacoEconomics*, *24*(12), 1199–1220.
- Rosner, B. (2006). *Fundamentals of biostatistics* (6th ed.). Belmont: Thomson Brooks/Cole.
- Roy, K. M., Roberts, M. C., & Canter, K. S. (2013). Anomalies in measuring health-related quality of life: using the PedsQL 4.0™ with children who have peanut allergy and their parents. *Applied Research in Quality of Life*, *8*, 511–518.
- Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment*, *26*(1), 1–2.
- Sheffler, L. C., Hanley, C., Bagley, A., Molitor, F., & James, M. A. (2009). Comparison of self-reports and parent proxy-reports of function and quality of life of children with below-the-elbow deficiency. *Journal of Bone and Joint Surgery*, *91*, 2852–2859.
- Su, C.-T., Wang, J.-D., & Lin, C.-Y. (2013). Child-rated versus parent-rated quality of life of community-based obese children across gender and grade. *Health and Quality of Life Outcomes*, *11*, 206.
- Su, C.-T., Ng, H.-S., Yang, A.-L., & Lin, C.-Y. (2014). Psychometric evaluation of the Short Form 36 Health Survey (SF-36) and the World Health Organization Quality of Life Scale Brief Version (WHOQOL-BREF) for patients with schizophrenia. *Psychological Assessment*, *26*(3), 980–989.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*, 469–477.
- Tomás, J. M., Hontangas, P. M., & Oliver, A. (2000). Linear confirmatory factor models to evaluate multitrait-multimethod matrices: the effects of number of indicators and correlation among methods. *Multivariate Behavioral Research*, *35*(4), 469–499.
- Trandis, H. C. (1995). *Individualism & collectivism*. Boulder: Westview Press.
- 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.
- Upton, P., Lawford, J., & Eiser, C. (2008). Parent-child agreement across child health-related quality of life instruments: a review of the literature. *Quality of Life Research*, *17*(6), 895–913.

- Varni, J. W., Seid, M., & Rode, C. A. (1999). The PedsQL: measurement model for the pediatric quality of life inventory. *Medical Care*, *37*(2), 126–139.
- 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*(8), 800–812.
- Varni, J. W., Limbers, C. A., & Newman, D. A. (2008). Factorial invariance of the PedsQLTM 4.0 generic core scales 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.
- World Health Organization. (1993). Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL). *Quality of Life Research*, *2*(2), 153–159.
- Young, D., Limbers, C. A., & Grimes, G. R. (2013). Is body mass index or percent body fat a stronger predictor of health-related quality of life in rural Hispanic youth? *Applied Research in Quality of Life*, *8*, 519–529.