

# Cross-Cultural Measurement Invariance of Adolescent Self- Report on the Pediatric Quality of Life Inventory™ 4.0

---

Stevanović, Dejan; Atilola, Olayinka; Vostanis, P.; Pal Singh Balhara, Y; Avicenna, Mohamad; Kandemir, H.; Knez, Rajna; Franić, Tomislav; Petrov, P.; Maroco, J.; ...

Source / Izvornik: **Journal of Adolescent Research**, 2016, 26, 687 - 695

Journal article, Published version

Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

<https://doi.org/10.1111/jora.12218>

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:184:475624>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-12-25**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Medicine - FMRI Repository](#)



## Cross-Cultural Measurement Invariance of Adolescent Self-Report on the Pediatric Quality of Life Inventory™ 4.0

Dejan Stevanovic

*Clinic for Neurology and Psychiatry for Children and Youth*

Olayinka Atilola

*Lagos State University College of Medicine Ikeja*

Panos Vostanis

*Leicester University*

Yatan Pal Singh Balhara

*All India Institute of Medical Sciences*

Mohamad Avicenna

*State Islamic University Syarif Hidayatullah*

Hasan Kandemir

*Harran University*

Rajna Knez

*University Hospital Centre Rijeka*

Tomislav Franic

*University of Split*

Petar Petrov

*University Hospital St. Marina*

João Maroco

*ISPA-Instituto Universitário*

Zorica Terzic Supic

*University of Belgrade*

Zahra Bagheri

*Shiraz University of Medical Sciences*

This study evaluated the cross-cultural measurement invariance of the Pediatric Quality of Life Inventory version 4.0 (PedsQL™) among adolescents sampled from Bulgaria, Croatia, India, Indonesia, Nigeria, Serbia, and Turkey. The multiple-indicator multiple-cause (MIMIC) model was used, which allowed controlling of demographic variables (i.e., age, gender, and socioeconomic status). Significant effects of country on scores within the PedsQL™ domains were observed, with up to 17 items showing differential item functioning (DIF) across the countries. We did not find support for cross-cultural measurement invariance hypotheses for scores on the PedsQL™ adolescent self-report in this study. Researchers should use caution in making cross-cultural quality of life comparisons while using the PedsQL.

Research suggests that the health-related quality of life (QOL) construct is strongly influenced by values, traditions, and beliefs about health in one's culture (e.g., Acquadro et al., 2003; Bullinger, 1997; Hutchinson, 1996; Schmidt & Bullinger, 2003; Stewart & Napoles-Springer, 2000). As such, QOL questionnaires are inherently sensitive to the language, dialect, customs, beliefs, and traditions of the cultures where they are developed (Sartorius & Kuyken, 1994). Administering a questionnaire to individuals outside of its intended culture would require translations and cultural adaptations to accommodate for the inherent linguistic (e.g., semantics) or cultural differences (e.g., religion;

Beaton, Bombardier, Guillemin, & Ferraz, 2000). A prerequisite for cross-cultural comparisons when using a translated QOL questionnaire is that the same theoretical construct is measured in each language/culture in the same way. In other words, construct invariance must be demonstrated for that questionnaire when tested simultaneously across several languages/cultures (Horn & McArdle, 1992). This is known as cross-cultural measurement invariance. When a questionnaire lacks cross-cultural measurement invariance, differences in means and other estimates observed across countries cannot be relied upon as true differences (e.g., Byrne & Watkins, 2003; Gregorich, 2006).

It is widely accepted that satisfactory cross-cultural measurement equivalence for a questionnaire

---

Funding source: None.

Requests for reprints should be sent to Dejan Stevanovic, Clinic for Neurology and Psychiatry for Children and Youth Belgrade, Dr. Subotic 6a, 11000 Belgrade, Serbia.  
E-mail: dejanstevanovic@eunet.rs

© 2015 The Authors

Journal of Research on Adolescence © 2015 Society for Research on Adolescence  
DOI: 10.1111/jora.12218

is demonstrated through linguistic, conceptual, metric, and functional equivalence (Johnson, 1998). Within a multicultural context, linguistic equivalence refers to the linguistic accuracy of items. Conceptual equivalence refers to the sameness in the meaning of factors/concepts. Metric equivalence refers to the sameness in psychometric properties. Functional equivalence refers to the fact that two or more behaviors in different cultures are functionally related to the same problem (Johnson, 1998). It is possible to examine aspects of linguistic and conceptual equivalence through qualitative methods that directly evaluate item and scale meanings during the translation and cultural adaptation process. A quantitative examination of metric equivalence for a questionnaire is conducted by administering the questionnaire simultaneously across several cultural groups, and using various methods within structural equation modeling (SEM) or in the item response theory (IRT) frameworks (e.g., Brown, 2006; Byrne & Watkins, 2003; Raju, Laffitte, & Byrne, 2002).

The Pediatric Quality of Life Inventory™ version 4.0 (PedsQL™) is one of the most frequently used generic QOL questionnaires among adolescents in multinational contexts (Varni, Seid, & Kurtin, 2001). It was developed in American English and has been translated and culturally adapted to over 70 languages and cultures (for details see <http://www.pedsq.org/>). It was initially proposed that the underlying measurement model for the PedsQL™ has five factors across its four scales: Physical Functioning, Emotional Functioning, Social Functioning, and School Functioning split into two different factors (Varni et al., 2001). This five-factor model was supported in later factor analytic studies with the original version (Limbers, Newman, & Varni, 2008a,b,c), as well as with different language versions among Norwegian (Reinfjell, Diseth, Veenstra, & Vikan, 2006), Greek (Gkoltsiou et al., 2008), Serbian (Stevanović, Lakić, & Damnjanović, 2011), and Iranian (Amiri et al., 2012) samples. Other studies employing Korean (Kook & Varni, 2008), Swedish (Petersen, Hagglof, Stenlund, & Bergstrom, 2009), Chinese (Hao, Tian, Lu, Chai, & Rao, 2010), Serbian (Stevanović et al., 2011), and Yoruba (Atilola & Stevanović, 2014) samples provided support for a four-factor model with each factor corresponding to one PedsQL™ scale, with the School Functioning scale being a single rather than a two-factor scale.

These findings indicate that the factorial structure of the PedsQL™ scale is inconsistent across the different language versions. As such, there is a

need to specifically examine the cross-cultural invariance of the scale to determine whether it produces comparable data across different cultures using different language versions. Several studies have reported the measurement invariance of the PedsQL™ factor model across age, gender, health status, or socioeconomic status using the original U.S. (Limbers et al., 2008a,b,c; Varni, Limbers, & Newman, 2008), Chinese (Lin et al., 2012), and Swedish (Petersen et al., 2009) self-report versions. However, very few studies have focused on cross-cultural measurement invariance of the PedsQL™.

In one study, the authors used multigroup confirmatory factor analysis (MG-CFA) to demonstrate measurement invariance across four race/ethnic groups in the United States (i.e., white non-Hispanic, Hispanic, Asian/Pacific Islander, and black non-Hispanic; Limbers, Newman, & Varni, 2009), indicating that scores on the PedsQL™ can produce comparable QOL measurements across these groups. Another study also used MG-CFA findings and supported the measurement invariance of the PedsQL™ across English- and Spanish-language groups in a Hispanic population in California (Newman, Limbers, & Varni, 2010). A major limitation to these studies was that their samples were limited to ethnic groups within the same country, and it is unclear whether these findings can be generalized to people of different ethnicities in their country of origin, where they are not an ethnic minority (Limbers et al., 2009; Newman et al., 2010).

This study evaluated the cross-cultural measurement invariance of the PedsQL™ by gathering data across seven countries that varied in terms of their stages of economic development, culture, and religious backgrounds. Specifically, the PedsQL™ was completed by adolescents living in Bulgaria, Croatia, India, Indonesia, Nigeria, Serbia, and Turkey as part of a project organized by the International Child Mental Health Study Group (ICMH-Study Group; Atilola, Balhara, Stevanovic, Avicenna, & Kandemir 2013).

## METHOD

### Participants

Data were gathered from 2,367 adolescents aged 13–18 years from Bulgaria, Croatia, India, Indonesia, Nigeria, Serbia, and Turkey. Demographic data on gender, age, and socioeconomic status were collected. A summary of these demographic data is provided in Table 1.

TABLE 1  
Distribution of Participants by Gender, Age, and Socioeconomic Status Across Seven Countries

Country	Gender*, Male/Female, n (%)	Age, M (SD) Years**	FAS Score, Median (Range)
Bulgaria, n = 265	129 (48.7)/136 (51.3)	15.33 (1.11)	6 (0–9)
Croatia, n = 293	121 (41.3)/172 (58.7)	16.19 (1.23)	6 (1–9)
India, n = 393	244 (62.1)/149 (37.9)	14.60 (0.68)	4 (2–5)
Indonesia, n = 228	105 (46.1)/123 (53.9)	16.13 (0.76)	6 (3–9)
Nigeria, n = 522	236 (45.2)/286 (54.8)	14.98 (1.26)	3 (0–8)
Serbia, n = 386	173 (44.8)/213 (52.2)	16.68 (1.02)	4 (0–9)
Turkey, n = 280	176 (62.9)/104 (37.1)	16.16 (0.95)	3 (0–8)

\* $\chi^2$  (6) = 60.89,  $p < .001$ ; \*\* $F$  (6) = 201.73,  $p < .0001$ .

Participants represented a sample of convenience from rural and urban communities based on the location of the researchers, and the same recruitment procedure was followed across all locations. The target sample per country was at least 560 adolescents in the 9th- to 12th-grade range. Samples from each location were drawn from at least two randomly selected postbasic schools, with at least one school from both a rural and an urban setting. Ethical approvals were obtained in all countries from the appropriate local authorities and/or ethical committees.

The adolescents were randomly contacted by school psychologists/counselors and were informed of the study. Of all contacted, only those who agreed to participate and returned the written self/parental consents (depending on age) were included. The response rate ranged from 45.6% to 93.2%. The questionnaires were administered to the adolescents while seated in school halls. They had enough space for comfort and privacy.

### PedsQL™ Questionnaire

The PedsQL™ consists of 23 items that are distributed across four scales: Physical Functioning (eight items); Emotional Functioning (five items); Social Functioning (five items); and School Functioning (five items) (Varni et al., 2001). All items have a 5-point response scale (0 = *never a problem*; 1 = *almost never a problem*; 2 = *sometimes a problem*; 3 = *often a problem*; and 4 = *almost always a problem*), which are reverse-scored and linearly transformed into a 0–100 scale, where higher scores indicate better functioning. Each scale score is computed as the sum of the items divided by the number of items answered on the scale. If more than 50% of the items on a scale are missing, a score is not computed. The Mapi Research Institute, which follows standardized procedures for cross-cultural adaptations of QOL questionnaires, provided the different

language versions of the PedsQL™ used in this study.

### Family Affluence Scale

Information about socioeconomic status (SES) was collected using the Family Affluence Scale (FAS), a valid indicator of adolescents' material circumstances (Boyce, Torsheim, Currie, & Zambon, 2006). The FAS is a self-report questionnaire that provides information about indicators of familial wealth using four parameters, including family car ownership, adolescent's own bedroom, family ownership of a computer, and family holiday in previous 12 months. A composite FAS score was calculated for each participant based on the responses to these items, with possible scores ranging 0–9. The greater the obtained score, the higher the SES.

### Data Analyses

As briefly reviewed in the Introduction, studies have suggested either a four- or five-factor model for the PedsQL™ factor structure. This study employed categorical confirmatory factor analysis (CCFA), which can appropriately model the ordered-categorical responses to assess for each country whether the four-factor model fit the data better than the five-factor one.

The multiple-indicator multiple-cause (MIMIC) model, which includes the factor model and additional exogenous variables, was used to assess cross-cultural measurement invariance across countries (Joreskog & Goldberger, 1975). This model consisted of a measurement component and a structural component. At the structural level, the latent variables are regressed onto covariate grouping variables such as country, while at the measurement level, the latent variables are each represented by observed indicators (i.e., the items). Measurement noninvariance is indicated when a

grouping variable has a statistically significant effect directly on an indicator, unmediated by the latent factor. The MIMIC enables uniform differential item functioning (DIF; i.e., scalar invariance) to be examined. An item is considered to display uniform DIF when people from different groups have different probabilities in item responses, despite having the same underlying level of a latent trait (Green, 1994). MIMIC also has the advantage of allowing control of covariate effects when testing measurement invariance across a specific grouping variable (e.g., Bye, Gallicchio, & Dykacz, 1985; Muthén, 1989). As age, gender, and SES differed significantly across the seven countries in our study, we utilized the advantage of this model to assess measurement invariance across the countries while controlling the effect of above-mentioned covariates.

In our study, all factors were simultaneously regressed on age and SES (as continuous variables) as well as the dummy variables of gender and country. Six binary indicators were defined for country, with one reference group. To conduct all pairwise comparison across the studied countries, each country was used in turn as the reference one, resulting in seven MIMIC analyses. In the first step, we tested the fit of a MIMIC model in which all possible direct effects from items to covariates were constrained to zero. If this model fits the data well, it would suggest an absence of noninvariant items; conversely, a poor-fitting model would indicate sources of noninvariance. Large modification indices would indicate which parameters are likely sources of noninvariance, and each of these parameters may then be estimated one at a time (Bye et al., 1985).

The fit of invariant and noninvariant models was evaluated using several criteria, including chi-square statistics, root mean square error of approximation (RMSEA), Tucker–Lewis index (TLI), and comparative fit index (CFI). CFI and TLI values  $\geq .95$ , and an RMSEA  $\leq .08$  suggest acceptable model fit. We applied the mean- and variance-adjusted weighted least square (WLSMV) estimation procedure which is recommended for ordinal indicators in the Mplus 6.1 software (Muthén & Muthén, 1998–2010) to fit CCFA and MIMIC models.

## RESULTS

Data from 2,367 adolescents were available for all analyses. There were statistically significant differences between the countries in age ( $p < .001$ ) and

gender ( $p < .001$ ), as well as differences in SES as indicated by different medians of the FAS scores (Table 1).

As seen in Table 2, there were trivial differences in the values of fit indices between the four-factor and five-factor models across each country. Considering that the PedsQL™ was originally proposed to have four factors, each represented by a scale, the four-factor model was used in the subsequent analyses. India was used as the initial reference country as its data provided the best model fit.

To assess measurement invariance of the PedsQL™ across countries while controlling the effect of age, gender, and SES as confounding variables, the initial model specified all latent variables to regress onto age, sex, SES, and country simultaneously, and all direct effects from each covariate to individual items were constrained to zero (i.e., no-DIF model). The values of fit indices ( $\chi^2(395) = 4320.65$ , CFI = .86, TLI = .84, RMSEA = .08) showed that this model did not fit to the data adequately, indicating the presence of DIF items. Modification indices suggested several sources of DIF, including a number of paths from countries to 13 items and from SES to three items. Each of these paths was examined one at a time through fitting several MIMIC models. Item parameter estimates for the final model in which 13 items were considered as uniform DIF (or had significant direct effects) are presented in Table 3. This model fitted adequately to the data ( $\chi^2(359) = 1939.89$ , CFI = .96, TLI = .95, RMSEA = .04). As uniform DIF is equivalent to the lack of consistency of threshold parameters (i.e., intercept in the MIMIC model), the value of  $\gamma$  coefficients in Table 3 indicates whether a given item is harder or easier for a specific country compared to the reference country. For instance,

TABLE 2  
Goodness of Fit Statistics for the Four-factor and Five-factor Models of the PedsQL™

	Four-Factor Model			Five-Factor Model		
	RMSEA	CFI	TLI	RMSEA	CFI	TLI
India	.05	.94	.93	.04	.96	.95
Serbia	.08	.88	.87	.07	.89	.87
Nigeria	.06	.94	.93	.05	.95	.94
Turkey	.13	.79	.76	.13	.79	.76
Indonesia	.08	.77	.78	.06	.93	.92
Bulgaria	.06	.92	.91	.06	.93	.92
Croatia	.09	.89	.87	.07	.91	.89

Note. RMSEA = root mean square error of approximation; TLI = Tucker–Lewis index; CFI = comparative fit index.

TABLE 3  
Results for the Multiple-Indicator Multiple-Cause (MIMIC) Models When India Was Considered as the Reference Country

Factor/Item	Age $\gamma$ (SE)	Sex $\gamma$ (SE)	SES $\gamma$ (SE)	Serbia $\gamma$ (SE)	Nigeria $\gamma$ (SE)	Turkey $\gamma$ (SE)	Indonesia $\gamma$ (SE)	Bulgaria $\gamma$ (SE)	Croatia $\gamma$ (SE)
Physical Functioning	.05 (.02)*	-.02 (.01)*	-.21 (.03)*	.29 (.06)*	-.002 (.05)	.13 (.06)*	.35 (.06)*	-.06 (.06)	.31 (.06)*
P1. Hard to walk	—	—	—	—	.48 (.08)*	—	—	—	—
P2. Hard to run	—	—	—	—	—	—	—	—	—
P3. Hard to do sports	—	—	—	—	—	—	—	—	—
P4. Hard to lift heavy things	—	—	.35 (.07)*	—	.96 (.13)*	1.11 (.14)*	—	—	-.38 (.09)*
P5. Hard to take bath/shower	—	—	.25 (.05)*	—	—	.46 (.10)*	—	—	—
P6. Hard to do chores around house	—	—	—	—	—	—	—	—	—
P7. Hurt or aches	—	—	—	-.31 (.09)*	—	—	—	—	—
P8. Low energy	—	—	—	—	—	—	—	—	—
Emotional Functioning	.05 (.02)*	-.05 (.01)*	-.34 (.04)*	.08 (.07)	.01 (.06)	-.26 (.07)*	.30 (.08)*	.11 (.07)	.15 (.37)*
P9. Feel afraid or scared	—	—	—	—	—	—	—	—	—
P10. Feel sad or blue	—	—	—	—	-.59 (.07)*	.34 (.10)*	—	—	—
P11. Feel angry	—	—	—	-.56 (.09)*	-.93 (.07)*	-.24 (.10)*	-.79 (.09)*	-.48 (.08)*	—
P12. Trouble sleeping	—	—	.24 (.05)*	.57 (.09)*	—	.97 (.09)*	—	—	—
P13. Worry	—	—	—	—	—	.62 (.09)*	—	—	—
Social Functioning	.05 (.02)*	-.03 (.01)*	.06 (.04)	-.18 (.07)*	.14 (.06)*	-.18 (.08)*	.43 (.08)*	.09 (.078)	.09 (.08)
P14. Trouble getting along w/peers	—	—	—	—	—	—	—	—	—
P15. Kids not wanting to be friend	—	—	—	—	—	—	—	—	—
P16. Teased	—	—	—	—	.08 (.08)*	-.25 (.09)*	—	.54 (.102)*	—
P17. Doing things other peers do	—	—	—	—	—	—	—	—	—
P18. Hard to keep up with others	—	—	—	-.54 (.11)*	—	.46 (.11)*	—	—	—
School Functioning	.10 (.02)*	-.02 (.01)*	.01 (.04)	.02 (.07)	-.38 (.06)*	.51 (.09)*	.42 (.08)*	-.07 (.08)	-.06 (.08)
P19. Hard to concentrate	—	—	—	—	-.36 (.06)*	-.93 (.10)*	—	—	—
P20. Forget things	—	—	—	—	—	-.61 (.10)*	—	—	—
P21. Trouble keeping up schoolwork	—	—	—	—	—	—	—	—	—
P22. Miss school—not well	—	—	—	—	—	—	—	—	—
P23. Miss school—doctor appointment	—	—	—	—	—	-.83 (.09)*	—	.58 (.09)*	.72 (.09)*

\* $p < .001$ ,  $\gamma$ , unstandardized coefficient; SE = standard error.



Serbian children reported fewer problems for item 7 in the Physical Functioning, item 11 in the Emotional Functioning, and item 18 in the Social Functioning domains compared to India as the reference country, indicated by the negative value of  $\gamma$  coefficients ( $-.31$ ,  $-.55$ , and  $-.54$ , respectively). In contrast, Serbian children reported more problems for item 12 in the Emotional Functioning domain as opposed to Indian children due to the positive value of  $\gamma$  coefficients (.57). Similar comparisons can be observed among the other countries included.

To carry out all pairwise comparisons, each country was in turn used as the reference country. Due to space limitations, this article selectively discusses some of the important aspects of the results, while other results are available upon request. When we considered Serbia as the reference country, nine uniform DIF items were detected, five of which (# 5, 8, 18, 19, and 20) exhibited DIF across Serbia and Turkey. Serbian adolescents reported more problems in items 5, 8, and 18 than Turkish children and fewer problems for items 19 and 20. Moreover, four DIF items were detected across Serbia and Nigeria; two of them were in the favor of Serbian adolescents and two of them were in the favor of Nigerian. Item 18 exhibited DIF across Serbia and Indonesia as well as items 11 and 12 across Serbia and India. When Nigeria was considered as the reference country, 11 items were detected as uniform DIF. Five items (# 8, 9, 12, 22, and 23) were noninvariant across Nigeria and Turkey, 4 items (# 1, 5, 9, and 13) across Nigeria and Croatia, 3 items (# 5, 11, and 13) across Nigeria and India, and 3 items (#5, 14, and 19) across Serbia and Nigeria. Nigerian adolescents reported more problems for most of the items compared with other countries. When we considered Turkey, Indonesia, Bulgaria, and Croatia as the reference categories, similar items were flagged with DIF; items 5, 8, 9, 11, 19, 20, and 23 were detected as DIF in all MIMIC models. For instance, when Turkey was considered as the reference country, Turkish adolescents reported fewer problems when endorsing items 5, 8, 12, 13, 22, and 23, and more problems for items 9, 16, 19, and 20. However, Indonesian adolescents reported more problems in almost all DIF items (# 5, 11, and 23) as opposed to Indian, Nigerian, and Turkish adolescents, except for item 18 as compared with Serbian adolescents. Moreover, Bulgarian adolescents reported fewer problems when endorsing items 16, 19, and 20 compared to Turkish, Indonesian, and Indian adolescents, and more problems for item 18 than Turk-

ish and Indonesian adolescents. Finally, when Croatia was the reference country, its adolescents reported more problems when endorsing items 1, 5, and 9 than Serbian, Nigerian, and Turkish adolescents and fewer problems for items 19, 20, and 23 than Turkish, Indonesian, and Nigerian adolescents.

Finally, the results indicated that age and sex had significant effects on all four QOL domains, but no DIF items were detected for these two demographic variables. In addition, SES had significant effects on the Physical and Emotional Functioning domains but not on the Social and School Functioning domains. Regression coefficients for items 5, 6, and 12 on SES implied that these items were identified with uniform DIF.

## DISCUSSION

This is the only study designed to evaluate the cross-cultural measurement invariance of the PedsQL™ adolescent self-report across several countries with different socioeconomic, cultural, and religious backgrounds. The results indicated that the PedsQL™ measurement model using the four domains represented by the PedsQL™ scales demonstrated cross-cultural measurement noninvariance. In other words, the PedsQL™ did not demonstrate equivalency in measuring QOL across different countries.

The initial CCFA showed that the four-factor and five-factor models for the PedsQL™ had different fit values across the countries included in the study. Fit indices from only three of the seven countries approached acceptable levels, namely Bulgaria, India, and Nigeria. There were only slight differences in the fit indices between the four- and five-factor models. Generally, results from the current study are consistent with previous studies that provided mixed findings regarding the measurement model for the PedsQL™ (e.g., Atilola & Stevanovic, 2014; Hao et al., 2010; Kook & Varni, 2008; Petersen et al., 2009; Stevanović et al., 2011).

In terms of the effect of demographic variables, the findings indicated that age and sex had significant effects on all four PedsQL™ domains. Additionally, there were significant effects of SES on the Physical and Emotional Functioning domains, with only three items exhibiting DIF. In general, these findings are in line with several previous studies showing that QOL is dependent on age and gender or SES (e.g., Michel, Bisegger, Fuhr, & Abel, 2009; Von Rueden, Gosch, Rajmil, Bisegger, & Ravens-Sieberer, 2006).

Because there were very small differences in the level of fit between the four- and five-factor measurement models across the countries, the MIMIC model was tested using the four-factor model, which was considered more parsimonious as well as more consistent with the four-scale structure of the PedsQL. The main analyses showed that the four-factor model is cross-culturally noninvariant and that there are some cultural influences or specific country traits in perceiving QOL domains as measured by the PedsQL™. Our results contradict the findings of a previous study that supported the measurement invariance of the PedsQL™ across English- and Spanish-language groups in a Hispanic population (Newman et al., 2010). In that study, the authors found only 6 non-invariant items of 23. In our study, it was observed that the number of DIF items varies depending on which country was selected as the reference. Overall, 17 different items exhibited DIF, with another six to 13 DIF items depending on the reference country. For example, 13 DIF items were identified when India was the reference country, while nine were identified when Serbia was used as reference. This finding implies that DIF occurs for 17 items at all levels along the latent trait, implying that adolescents across the countries endorsed these items differently (Green, 1994). Considering that we controlled for effects of age, gender, and SES, our findings indicated that PedsQL™ items exhibiting DIF are more sensitive and easily confounded by the culture-specific attributes related to the construct.

The fact that a majority of the PedsQL™ items exhibited DIF limits the intercountry comparisons that can be made from this tool's QOL dimensions. Our findings suggest that the norms for a particular dimension in one culture could confound cross-cultural comparisons. It is possible that the different culture-specific attributes operating in one country relate to the QOL constructs differently, insofar as the DIF items may not adequately capture or represent an area of QOL that is important or relevant to that country's cultural norms (Heine, Lehman, Peng, & Greenholtz, 2002). This is best recognized by the translation and cultural adaptation of a questionnaire into other languages (Berry, Poortinga, Segall, & Dasen, 2002). Cultural adaptation of a questionnaire is important for ensuring conceptual invariance in measurement to avoid over- or underevaluations of a construct from different ethnic groups (Berry et al., 2002). Considering that the PedsQL™ has a standardized approach to translation and cultural adaptation to ensure conceptual invariance, it is likely that the relevance

of some items varied among adolescents from different cultures due to cultural norms.

Our findings have several important research implications. The findings suggest that the current PedsQL™ self-report measurement model using four scales does not allow cross-cultural comparisons in levels of adolescent QOL. However, this does not mean that the PedsQL™ should not be used for within-country comparisons, especially when country-specific norms are available. In fact, given that there is support for the measurement invariance of the PedsQL™ across different cultural background within the same country (Newman et al., 2010), our findings highlight the possibility that the instrument may be more useful for within-country than for between-country cultural comparisons. If PedsQL™ items with DIF are taken into account in between-country comparisons of QOL, it may be possible to use the questionnaire, but more research is warranted in this regard.

The current study had several strengths. The current study utilized a large sample size, used multiple language versions of the PedsQL™, employed a cross-country design, and included socioeconomically, culturally, and religiously diverse nations. All these elements allowed us to cover greater variability among adolescents and countries and to broaden the generalizability of findings to a multicultural context. There were, however, some limitations to the present study that need to be taken into consideration when interpreting our findings. First, participants were sampled from regions of convenience. Although schools in the regions were randomly selected, using a sample of convenience could limit generalizability of the findings to adolescents from other country's regions. Second, the response rate varied from very low (45.6%) in Indonesia to very high (93.2%) in Nigeria, which could further limit the generalizability of the findings, given that the reasons behind such differences in response rates are unclear. An example may be that some students were more concerned about the confidentiality of their responses from school personnel, despite the fact that the questionnaires were returned sealed. Third, it has been argued that a wide range in sample size could bias the results when conducting MG-CFA (Brown, 2006, p. 279), which might be possible with our study due to a varied sample size across the included countries. Fourth, considering that we only gathered adolescents' self-report responses, there might be less noninvariance if child and/or parent reports were used, which could be explored in future studies. Fifth, it might



be more fruitful to use IRT methods as alternatives to testing measurement invariance. Finally in the present study, the effect of clustered data has not been taken into account through multilevel structural equation modeling, because at least 20 to 50 clusters are needed to ensure stable parameter estimates and convergence in a model fitting (Stegmueller, 2013). As the present study included only seven countries, we could not utilize the advantage of multilevel structure equation modeling such as multilevel MIMIC or CFA.

In summary, we did not find support for cross-cultural measurement invariance hypotheses for the four scales of the PedsQL™ adolescent self-report in this study. Researchers should use caution in making cross-cultural QOL comparisons using the published PedsQL™ scales until further research is conducted.

## REFERENCES

- Acquadro, C., Berzon, R., Dubois, D., Leidy, N. K., Marquis, P., Revicki, D., Rothman, M., & The PRO Harmonization Group. (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.
- Amiri, P., Eslamian, G., Mirmiran, P., Shiva, N., Jafarabadi, M. A., & Azizi, F. (2012). Validity and reliability of the Iranian version of the Pediatric Quality of Life Inventory™ 4.0 (PedsQL™) Generic Core Scales in children. *Health and Quality of Life Outcomes, 10*, 1–9. doi:10.1186/1477-7525-10-3
- Atilola, O., Balhara, Y. P. S., Stevanovic, D., Avicenna, M., & Kandemir, H. (2013). Self-reported mental health problems among adolescents in developing countries: Results from an international pilot sample. *Journal of Developmental and Behavioral Pediatrics, 34*, 129–137. doi:10.1097/DBP.0b013e31828123a6
- Atilola, O., & Stevanović, D. (2014). PedsQL™ 4.0 Generic Core Scales for adolescents in the Yoruba language: Translation and general psychometric properties. *Clinical Child Psychology and Psychiatry, 19*, 286–298. doi:10.1177/1359104513488375
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*, 3186–3191.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). Methodological concerns. In J. W. Berry, Y. H. Poortinga, M. H. Segall, & P. R. Dasen (Eds.), *Cross-cultural psychology: Research and applications* (pp. 286–316). Cambridge, UK: Cambridge University Press.
- Boyce, W., Torsheim, T., Currie, C., & Zambon, A. (2006). The family affluence scale as a measure of national wealth: Validation of an adolescent self-report measure. *Social Indicators Research, 78*, 473–487. doi:10.1007/s11205-005-1607-6
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Bullinger, M. (1997). The challenge of cross-cultural quality of life assessment. *Psychology and Health, 12*, 815–825. doi:10.1080/08870449708406742
- Bye, B. V., Gallicchio, S. J., & Dykacz, J. M. (1985). Multiple-indicator, multiple-cause models for a single latent variable with ordinal indicators. *Sociological Methods and Research, 13*, 487–509. doi:10.1177/0049124185013004003
- Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*, 155–175. doi:10.1177/0022022102250225
- Gkoltsiou, K., Dimitrakaki, C., Tzavara, C., Papaevangelou, V., Varni, J. W., & Tountas, Y. (2008). Measuring health-related quality of life in Greek children: Psychometric properties of the Greek version of the Pediatric Quality of Life Inventory™ 4.0 Generic Core Scales. *Quality of Life Research, 17*, 299–305. doi:10.1007/s11136-007-9294-1
- Green, B. F. (1994). Differential item functioning: Techniques, findings, and prospects. In D. Laveault, B. D. Zumbo, M. E. Gessaroli, & M. W. Boss (Eds.), *Modern theories of measurement: Problems and issues*. Ottawa, Canada: University of Ottawa Press.
- 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*(11 Suppl. 3), S78. doi:10.1097/01.mlr.0000245454.12228.8f
- Hao, Y., Tian, Q., Lu, Y., Chai, Y., & Rao, S. (2010). Psychometric properties of the Chinese version of the Pediatric Quality of Life Inventory™ 4.0 generic core scales. *Quality of Life Research, 19*, 1229–1233. doi:10.1097/NCC.0b013e31827028c8
- Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology, 82*, 903. doi:10.1037//0022-3514.82.6.903
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117–144. doi:10.1080/03610739208253916
- Hutchinson, J. F. (1996). Quality of life in ethnic groups. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (pp. 587–593). Philadelphia, PA: Lippincott-Raven.
- Johnson, T. P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA-NachrichtenSpezial, 3*, 1–40.
- Joreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*, 631–639. doi:10.2307/2285946

- Kook, S. H., & Varni, J. W. (2008). Validation of the Korean version of the pediatric quality of life inventory 4.0 (PedsQL) generic core scales in school children and adolescents using the Rasch model. *Health and Quality of Life Outcomes*, *6*, 41. doi:10.1186/1477-7525-6-41
- Limbers, C. A., Newman, D. A., & Varni, J. W. (2008a). Factorial invariance of child self-report across healthy and chronic health condition groups: A confirmatory factor analysis utilizing the PedsQL™ 4.0 Generic Core Scales. *Journal of Pediatric Psychology*, *33*, 630–639. doi:10.1093/jpepsy/jsm131
- Limbers, C. A., Newman, D. A., & Varni, J. W. (2008b). Factorial invariance of child self-report across age subgroups: A confirmatory factor analysis of ages 5 to 16 years utilizing the PedsQL 4.0 Generic Core Scales. *Value in Health*, *11*, 659–668. doi:10.1111/j.1524-4733.2007.00289.x
- Limbers, C. A., Newman, D. A., & Varni, J. W. (2008c). Factorial invariance of child self-report across socioeconomic status groups: a multigroup confirmatory factor analysis utilizing the PedsQL™ 4.0 Generic Core Scales. *Journal of Behavioral Medicine*, *31*, 401–411. doi:10.1007/s10865-008-9166-3
- Limbers, C. A., Newman, D. A., & Varni, J. W. (2009). Factorial invariance of child self-report across race/ethnicity groups: A multigroup confirmatory factor analysis approach utilizing the PedsQL™ 4.0 Generic Core Scales. *Annals of Epidemiology*, *19*, 575–581. doi:10.1016/j.annepidem.2009.04.004
- Lin, C. Y., Luh, W. M., Cheng, C. P., Yang, A. L., Su, C. T., & Ma, H. I. (2012). Measurement Invariance across child self-reports and parent-proxy reports in the Chinese Version of the Pediatric Quality of Life Inventory Version 4.0. *Child Psychiatry & Human Development*, *44*, 583–590. doi:10.1007/s10578-012-0352-8
- Michel, G., Bisegger, C., Fuhr, D. C., & Abel, T. (2009). Age and gender differences in health-related quality of life of children and adolescents in Europe: A multi-level analysis. *Quality of Life Research*, *18*, 1147–1157. doi:10.1007/s11136-009-9538-3
- Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557–585. doi:10.1007/BF02296397
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.) Los Angeles, CA: Muthén & Muthén.
- Newman, D. A., Limbers, C. A., & Varni, J. W. (2010). Factorial invariance of child self-report across English and Spanish language groups in a Hispanic population utilizing the PedsQL™ 4.0 Generic Core Scales. *European Journal of Psychological Assessment*, *26*, 194–202. doi:10.1027/1015-5759/a000026
- Petersen, S., Hagglof, B., Stenlund, H., & Bergstrom, E. (2009). Psychometric properties of the Swedish PedsQL, pediatric quality of life inventory 4.0 Generic Core Scales. *Acta Paediatrica*, *98*, 1504–1512. doi:10.1111/j.1651-2227.2009.01360.x
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*, 517. doi:10.1037//0021-9010.87.3.517
- 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. doi:10.1186/1477-7525-4-61
- Sartorius, N., & Kuyken, W. (1994). Translation of health status instruments. In J. Orley & W. Kuyken (Eds.), *Quality of life assessment: International perspectives*. Berlin, Germany: Springer-Verlag.
- Schmidt, S., & Bullinger, M. (2003). Current issues in cross-cultural quality of life instrument development. *Archives of Physical Medicine and Rehabilitation*, *84*, S29–S34. doi:10.1053/apmr.2003.50244
- Stegmuller, D. (2013). How many countries for multi-level modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science*, *57*, 748–761. doi:10.1111/ajps.12001
- Stevanović, D., Lakić, A., & Damjanović, M. (2011). Some psychometric properties of the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales (PedsQL™) in the general Serbian population. *Quality of Life Research*, *20*, 945–949. doi:10.1007/s11136-010-9833-z
- Stewart, A. L., & Napoles-Springer, A. (2000). Health-related quality-of-life assessments in diverse population groups in the United States. *Medical Care*, *38*(9 Suppl), II102–II124.
- Varni, J., Limbers, C., & Newman, D. (2008). Factorial invariance of the PedsQL™ 4.0 Generic Core Scales child self-report across gender: A multigroup confirmatory factor analysis with 11,356 children ages 5 to 18. *Applied Research in Quality of Life*, *3*, 137–148. doi:10.1111/ajps.12001
- 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.
- 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