

Cross-cultural Measurement Equivalence of the KINDL Questionnaire for Quality of Life Assessment in Children and Adolescents

Peyman Jafari¹ · Dejan Stevanovic² · Zahra Bagheri¹

Published online: 17 July 2015

© Springer Science+Business Media New York 2015

Abstract This cross-cultural study aimed to assess whether Iranian and Serbian children, and also their parents, perceived the meaning of the items in the KINDL quality of life questionnaire consistently. The sample included 1086 Iranian and 756 Serbian children and adolescents, alongside 1061 and 618 of their parents, respectively. The ordinal logistic regression was used to assess differential item functioning (DIF) of the self and proxy-reports of the two versions of the KINDL, including Kid-KINDL and Kiddo-KINDL, across Iranian and Serbian samples. Statistically significant DIF was flagged for 14 out of 24 (58 %) and 20 out of 24 (83 %) items in the self-report of the Kid-KINDL and Kiddo-KINDL, respectively. Moreover, 20 out of 24 (83 %) in the proxy reports of the both Kid-KINDL and Kiddo-KINDL, showed DIF across two samples. Accordingly, considerable caution is warranted when using the KINDL for cross-cultural comparisons.

Keywords Quality of life · Differential item functioning · Cross-cultural · Children

Introduction

Health-related quality of life (HRQOL) is gaining significant research attention in terms of prevention, treatment and rehabilitation studies among different ethnic and

cultural groups worldwide [1]. However, studies have shown that the HRQOL concept is strongly influenced by cultural values, traditions, and beliefs [2–5]. Consequently, questionnaires designed for HRQOL assessment are inherently sensitive to the language, dialect and community values of the local cultures [6]. Hence, in cross-cultural research, not only should HRQOL instruments be translated well linguistically, but they also should be culturally adopted to preserve the content validity of the questionnaire across different cultures [3, 7]. As borrowed from cultural psychology, the primary requirement for the cross-cultural adaptation process of one questionnaire is to ensure its cross-cultural measurement equivalence [8]. The establishment of measurement equivalence is especially important in making valid comparisons of HRQOL scores across cultural groups. Measurement invariance evaluates whether the probability of responding to specific items within a measure is the same across the compared groups after controlling for the construct being measured [9]. For the quality of life (QOL) concept, this implies that the same theoretical construct is measured in two or more cultures in the same way [5]; and when differences in the QOL construct are found across different cultural or language groups, these differences should more likely be true difference between the groups rather than differences in the perception and interpretation of the items as a function of health status [10, 11].

To date, although many generic QOL questionnaires were developed for paediatric use, very few have documented cross-cultural measurement equivalence. For example, the KIDSCREEN is the first questionnaire with sound psychometric properties that was simultaneously developed across several European countries [12, 13]. Two recent studies provided evidence for the cross-cultural measurement equivalence of the KIDSCREEN using a

✉ Dejan Stevanovic
stevanovic.dejan79@gmail.com

¹ Department of Biostatistics, Shiraz University of Medical Sciences, Shiraz, Iran

² Department of Psychiatry, Clinic for Neurology and Psychiatry for Children and Youth, Dr. Subotic 6a, 11000 Belgrade, Serbia

short, self-report version, which was evaluated across multiple language versions [14, 15]. In addition, for the Pediatric Quality of Life Inventory Version 4.0TM (PedsQLTM 4.0) [16] self-report, the findings of a US study revealed that children and adolescents across four ethnic groups interpreted items and factors in a similar manner regardless of their ethnicity, which is indicative of cross-cultural measurement equivalence [11].

The KINDL [17], PedsQLTM 4.0 [16] and KIDSCREEN [12–14] are the most frequently used questionnaires in paediatric QOL studies. All three questionnaires are generic, multidimensional and originally designed to assess QOL in children and adolescents. However, a systematic review has shown that many paediatric instruments do not reflect QOL definitions provided by World Health Organization (WHO) [2]. By mapping the instruments to the International Classification of Functioning, Disability and Health for Children and Adolescents (ICF-CY), the researchers found that the same perspective of QOL cannot be obtained from each instrument [2]. The KIDSCREEN particularly measures HRQOL, the PedsQLTM 4.0 covers a wide definition of functioning, disability, and health (FDH), and the KINDL is an appropriate instrument to measure FDH with some HRQOL features. Therefore, researchers should be clear about whether they intend to measure HRQOL (which includes the expectations, standards, or concerns about health domains) or FDH (i.e., the performance, capacity, presence/absence, frequency and severity of psychosocial domains) [2].

The KINDL has previously been translated and culturally adapted for use in various languages and countries, and has shown acceptable psychometric properties [18–25]. Although translated forms of the KINDL may achieve linguistic equivalence (i.e., each translated word appropriately matches its translated counterpart), the additional aspect of equivalence (i.e., measurement equivalence) should be assessed to ensure that the KINDL operates exactly in the same way in different countries, especially in those with diverse cultural backgrounds.

Up to present, the measurement equivalence of the KINDL was only evaluated across Iranian children and their parents [21]. That study examined whether children and their parents responded consistently to the KINDL items, and it was found that the KINDL failed to produce equivalent measurement between the two groups [21]. Although a previous research has shown that the KIDSCREEN-27 can be used for cross-cultural comparison between Iranian and Serbian samples [15], such an explanation has never been provided to address whether the KINDL is an invariant measure for cross-cultural research. The aim of the present study was to evaluate the cross-cultural measurement equivalence of the KINDL using two

samples of Iranian and Serbian children and adolescents, alongside their parents. Iranian and Serbian samples were selected because of the followings. Iran and Serbia are two countries with developing economies, but the first is an Asian and the second is a European country. Moreover, there are significant religious and cultural differences between them. While children, adolescents, and their parents in Iran are mostly Muslims, their peers in Serbia are mostly Orthodox, which has led to the variations in the way of life, beliefs, traditions and laws and possibly different health perceptions. Finally, we would like to compare HRQOL scores across the two countries. However, if substantive DIF is evident, we follow a removing and retaining strategy to determine whether HRQOL scores, and the resultant conclusions, will be altered in a meaningful way with and without DIF items.

Methods

Participants

For the present study, data were used for children and adolescents aged eight to 16 from Serbia and Iran who participated in two previous studies [21, 23, 24]. In brief, the Iranian sample included 1086 school children and adolescents (62.4 % boys, 37.6 % girls) and 1061 parents, while the Serbian sample consisted of 756 children and adolescents (50.1 % boys, 49.1 % girls) and 618 parents. The mean (\pm SD) age of the Iranian and Serbian sample was 13.32 ± 2.26 and 12.23 ± 1.85 years, respectively. In both countries the samples were selected from urban areas. The Iranian sample was selected from Shiraz the largest city in southern of Iran and the Serbian sample was selected from Belgrade, the capital of Serbia. Moreover, almost all children in both cultures were healthy (more than 90 % of Iranians and 95 % of Serbians), and were recruited from public schools. In the present study, data from two separate studies [21, 23, 24], which were not originally designed for cross-cultural comparison, were merged to do a cross-cultural research. Hence, other sociodemographic variables including parents' income, education, age, and gender were not available in both samples.

While Iranian children, adolescents and their parents completed the questionnaire at home, almost all of the participants in the Serbian study completed the questionnaire in school. In both samples, children, adolescents and their parents signed the informed consent forms, and they were instructed in detail how to complete the KINDL. The process of translation and linguistic validation for the Iranian and Serbian version of the KINDL pertaining to other relevant details are fully described in [21, 23, 24].

Questionnaire

The KINDL [17] questionnaire includes a child self-report and parent proxy-report. Both KINDL reports contained 24 items in six subscales: physical well-being, emotional well-being, self-esteem, family, friends and school. There is a version for eight- to 12 year-old children (Kid-KINDL) and for 13- to 16 year-old adolescents (Kiddo-KINDL). In the present study, self and proxy-reports of the Kid-KINDL was completed by 353 and 345 of Iranian, and 356 and 282 of Serbian children and their parents, respectively. Self and proxy-reports of the Kiddo-KINDL was completed by 733 and 716 of Iranian, and 400 and 336 of Serbian adolescents and their parents, respectively. Both age versions had the same items and scoring format; the only difference was in the slightly different wording for an instance that was more relevant to one group than the other (i.e., item 17 in the child version is “played with friends”, while in the adolescent version, the same item is “doing things with friends”). All participants responded to the items on a 5-point Likert scale (0 = never, 1 = seldom, 2 = sometimes, 3 = often and 4 = all the time). For ease of interpretation, rating scale categories were reversed so that higher categories indicated better QOL. Overall raw subscale scores were formed from the item’s mean values. The raw sub-scale scores were transformed into a 0–100 scale, with higher scores indicating better QOL.

Statistical Analysis

In the present study, ordinal logistic regression (OLR) as a proportional odds model described by Swaminathan and Rogers (hereinafter referred as the SR criterion) was used to assess the measurement equivalence of the KINDL across Iranian and Serbian children and adolescents, alongside their parents [26, 27]. All 24 items are considered as the response variables in ordinal regression models. Testing for the presence of DIF (uniform and non-uniform) under the OLR model, adjusted by child’s gender, is based on comparing three different models as follows:

$$\text{Model 1 : Logit [P (Y} \leq \text{K)]} = \alpha_k + \beta_1 \times \theta + \gamma \times \text{gender}$$

$$\begin{aligned} \text{Model 2 : Logit [P (Y} \leq \text{K)]} \\ = \alpha_k + \beta_1 \times \theta + \beta_2 \times \text{country} + \gamma \times \text{gender} \end{aligned}$$

$$\begin{aligned} \text{Model 3 : Logit [P (Y} \leq \text{K)]} \\ = \alpha_k + \beta_1 \times \theta + \beta_2 \times \text{country} + \beta_3 \times \theta \\ \times \text{country} + \gamma \times \text{gender} \end{aligned}$$

The term θ is used here to represent the trait measured by the KINDL instrument as the observed sum score, and country is the grouping variable with two levels (Iran and Serbia). According to the above models, uniform DIF could

be detected by comparing the log likelihood values for Model 1 and 2, and non-uniform DIF by Model 2 and 3. For both uniform and non-uniform DIF twice the difference in log likelihoods is compared to a Chi-square distribution with one degree of freedom. According to SR criterion, a slight difference in log-likelihood of the Models for detecting uniform and non-uniform DIF could be statistically significant, given a large enough sample. Hence, in response to this concern, we used Zumbo and Gelin (ZG) and Crane, van Belle, and Larson (CvBL) criteria to quantify the magnitude of DIF, which may not be practically or clinically important. According to ZG criterion, the magnitude of uniform and non-uniform DIF could be determined by the difference in pseudo R^2 (ΔR^2) between Model 1 and 2, and Model 2 and 3, respectively. ΔR^2 values less than 0.035, between 0.035 and 0.070, and above 0.070 are classified as negligible, moderate, and large DIF, respectively. Moreover, according to CvBL criteria, the absolute proportion change in point estimates for β_1 from Models 1 and 2, $\Delta\beta_1 = |\beta_1(\text{model 1}) - \beta_1(\text{model 2})|/\beta_1(\text{model 1})$, is used to identify items with uniform DIF. Based on simulation studies, 10 % change in $\Delta\beta_1$ is considered as a practically meaningful effect. An advantage of the OLR method applied in the present study is that it provides effect size measures to quantify the magnitude of DIF, which may not be practically meaningful. In OLR approach, a slight difference in log-likelihood of models for testing uniform and non-uniform DIF could be statistically significant, given a large sample. Hence, inclusion of effect size reduces false discovery DIF detection rate [28]. In the present study, a removing and retaining strategy is used to assess the potential impact of DIF in comparing HRQOL scores across two countries [28].

Expected item score curves were also used to determine the direction and magnitude of DIF visually across Iranian and Serbian children and adolescents, alongside their parents. This curve is a function of θ and provides a better understanding of uniform and non-uniform DIF. It should be noted that for items with non-uniform DIF the direction of DIF differs along the subscale, leading to the effect of DIF cancel-out at the scale level [29]. Hence, expected item score curves are only depicted for items with uniform DIF. The OLR procedure was implemented using the computer program SAS 9.1.

Results

DIF Results for Kid-KINDL

The results of the OLR DIF analysis for the child self-report and parent proxy-report of the Kid-KINDL across

Iranian and Serbian samples are shown in Table 1. In the child self-reports of the Kid-KINDL, 14 out of 24 (58 %) items were flagged with DIF according to SR criterion. Of these 14 items, nine (64 %) items displayed uniform DIF and five (36 %) items displayed non-uniform DIF. However, based on the ZG criterion, five items exhibited uniform, where change in R-squared (ΔR^2) associated with these items ranged from 0.044 to 0.15. Moreover, according to the ZG criterion, item 1 in the school subscale exhibited non-uniform DIF ($\Delta R^2 = 0.187$). Additionally, four items had uniform DIF according to the CvBL criterion, where the change in β_1 coefficients ($\Delta\beta_1$) for these items ranged from 0.1 to 0.291. As shown in Fig. 1, item 3 in the physical subscale and item 2 in the emotional subscale had uniform DIF and their effects could not be cancelled-out at the domain level by uniform DIF items in the opposite direction. Moreover, for items 1, 3 and 4 with uniform DIF in the self-esteem subscale, item expected scores were in one direction and could not cancel one another out. In contrast, expected scores for items 2 and 4 in the family subscale, and items 1 and 2 in the friend subscale were in the opposite direction, indicating that the effect of uniform DIF could be cancelled-out for these subscales.

In the parent proxy-reports of the Kid-KINDL, 20 out of 24 (83 %) items were flagged with DIF according to SR criterion; 13 (65 %) items displayed uniform DIF and seven (35 %) items displayed non-uniform DIF. However, based on the ZG criterion, six items exhibited uniform DIF, where the difference in R-squared (ΔR^2) for these items varied from 0.036 to 0.173. According to the CvBL criterion six items had uniform DIF, where changes in β_1 coefficients associated with these items varied from 0.109 to 0.486. As shown in Fig. 2, of the three items with uniform DIF in the physical well-being subscale, the item expected scores for items 2 and 3 went in one direction, and item 4 went in the opposite direction, indicating that the effect of uniform DIF could not be cancelled-out for this subscale across the two samples. Similar pattern can also be observed in the emotional, self-esteem, and friend subscales. In contrast, there was just one item with uniform DIF in the family subscale; consequently, its effect could not be cancelled-out.

DIF Results for Kiddo-KINDL

The results of the OLR DIF analysis for the child self-report and parent proxy-report of the Kiddo-KINDL are shown in Table 2. In the child reports, 20 out of 24 (83 %) items were identified with DIF according to the SR criterion. Of these 20 items, 13 (65 %) items displayed uniform DIF and seven (35 %) items displayed non-uniform DIF across Iranian and Serbian adolescents. According to the

ZG criterion, items 1 and 2 in the self-esteem subscale displayed uniform DIF, where the change in R-squared associated with these items were 0.051 and 0.079, respectively. Moreover, according to the CvBL criterion item 1 in the self-esteem and item 2 in the friend subscales had uniform DIF. The changes in β_1 coefficients associated with these items were 0.211 and 0.114, respectively. As shown in Fig. 3, item expected scores for items 2 and 3 went in one direction, and item 1 went in the opposite direction, indicating that the effect of uniform DIF could not be cancelled-out for this subscale across Iranian and Serbian adolescents. Moreover, for items 1 and 3 with uniform DIF in the friend subscale, expected scores were in one direction and could not cancel one another out. Of the four items with uniform DIF in the physical well-being subscale, item 2 showed DIF in one direction, whereas items 1, 3 and 4 showed DIF in the opposite direction; hence, they could not be cancelled each other out. In contrast, items 2 and 3 in the family subscale were in the opposite direction, indicating that the effect of uniform DIF could be cancelled-out for this subscale. Moreover, in the friend and school subscales, only items 1 and 3 had uniform DIF, respectively; hence, their effects could not be cancelled-out at the scale level.

In the parent proxy-reports, the SR criterion showed that 20 out of 24 (83 %) items were flagged with DIF between Iranian and Serbian parents. Of these items, eight items (40 %) exhibited non-uniform DIF and 12 items (60 %) exhibited uniform DIF. According to the ZG criterion, five items exhibited uniform DIF, where the difference in R-squared (ΔR^2) for these items ranged from 0.042 to 0.155. Based on CvBL criterion, five items had uniform DIF, where changes in β_1 coefficients associated with these items varied from 0.106 to 0.545.

As shown in Fig. 4, in the physical subscale, item 3 showed DIF in one direction, whereas item 4 showed DIF in the opposite direction; hence, they cancelled each other out. In contrast, expected scores for items 1 and 2 in the emotional well-being subscale were in the same direction, and could not cancel each other out. Moreover, of the four items with the uniform DIF in the self-esteem subscale, items 1, 3, and 4 went in one direction, and item 2 went in the opposite direction, indicating that the effect of DIF cannot be cancelled-out at the scale level. Similarly, in the friend subscale, the item expected scores for items 2 and 3 went in one direction, and item 1 went in the opposite direction, indicating that the effect of uniform DIF could not be cancelled-out for this subscale across Iranian and Serbian parents. There was only one item with uniform DIF in the school subscale; consequently, its effect could not be cancelled-out at the scale level.

As shown in Table 3, Iranian children and their parents rated the child's QOL significantly lower than their Serbian

Table 1 The results of the ordinal logistic regression DIF analysis on the Kid-KINDL across Iranian and Serbian children (8–12)

Subscales	Non-uniform DIF		Uniform DIF		
	χ^2 ^a (P ^b)	ΔR^2 ^b	χ^2 (P)	ΔR^2	$\frac{\beta_1^* - \beta_1}{\beta_1}$
<i>Child self-reports</i>					
<i>Physical well-being</i>					
3. I was tired and worn-out	0.02 (0.87)	0.000	9.94 (0.002)	0.0084	0.003
<i>Emotional well-being</i>					
1. I had fun and laughed a lot	21.84 (<0.001)	0.023	1.81 (0.170)	0.002	0.005
2. I was bored	2.24 (0.13)	0.002	8.07 (<0.001)	0.007	0.018
3. I felt alone	4.95 (0.02)	0.004	0.891 (0.34)	0.000	0.000
<i>Self-esteem</i>					
1. I was proud of myself	1.11 (0.29)	0.008	128.3 (<0.001)	0.097	0.291
2. I felt on top of the world	5.55 (0.02)	0.003	218.7 (<0.001)	0.150	0.046
3. I felt pleased with myself	0.13 (0.71)	0.000	33.35 (<0.001)	0.023	0.100
4. I had lots of good ideas	2.09 (0.15)	0.002	27.74 (<0.001)	0.023	0.100
<i>Family</i>					
2. I felt fine at home	2.53 (0.11)	0.003	40.65 (<0.001)	0.044	0.006
3. We quarreled at home	4.87 (0.03)	0.005	20.62 (<0.001)	0.023	0.046
4. I felt restricted by my parents	2.04 (0.15)	0.002	4.81 (0.03)	0.004	0.013
<i>Friends</i>					
1. I did things together with my friends	1.07 (0.30)	0.001	88.52 (<0.001)	0.088	0.092
2. I was a “success” with my friends	0.11 (0.74)	0.000	158.52 (<0.001)	0.139	0.180
<i>School</i>					
1. doing the schoolwork was easy	19.57 (<0.001)	0.187	0.821 (0.36)	0.001	0.007
<i>Parent proxy-reports</i>					
<i>Physical well-being</i>					
2. My child had a headache or tummy ache	1.28 (0.25)	0.001	10.36 (0.001)	0.0098	0.029
3. My child was tired and worn-out	0.78 (0.37)	0.001	9.25 (0.002)	0.0083	0.022
4. My child felt strong and full of energy	2.82 (0.093)	0.003	7.56 (0.005)	0.007	0.006
<i>Emotional well-being</i>					
2. My child didn’t feel much like doing anything	0.002 (0.96)	0.000	39.44 (<0.001)	0.033	0.050
3. My child felt alone	0.828 (0.362)	0.001	13.22 (<0.001)	0.013	0.084
4. My child felt scared or unsure of him-/herself	0.004 (0.94)	0.000	33.25 (<0.001)	0.036	0.155
<i>Self-esteem</i>					
1. My child was proud of him-/herself	1.58 (0.21)	0.001	96.74 (<0.001)	0.070	0.208
2. My child felt on top of the world	0.46 (0.49)	0.000	77.15 (<0.001)	0.056	0.005
3. My child felt pleased with him-/herself	2.68 (0.10)	0.002	40.97 (<0.001)	0.032	0.109
4. My child had lots of good ideas	7.84 (0.005)	0.008	2.15 (0.142)	0.002	0.018
<i>Family</i>					
3. We quarreled at home	1.98 (0.15)	0.002	59.16 (<0.001)	0.057	0.187
4. My child felt that I was bossing him/her around	9.99 (0.002)	0.009	12.57 (<0.001)	0.012	0.030
<i>Friends</i>					
1. My child did things together with friends	2.31 (0.13)	0.002	140.91 (<0.001)	0.120	0.045
2. My child was liked by other kids	0.41 (0.52)	0.000	164.70 (<0.001)	0.173	0.486
3. My child got along well with his/her friends	1.38 (0.23)	0.001	24.90 (<0.001)	0.029	0.119
4. My child felt different from other children	8.54 (0.003)	0.009	0.135 (0.71)	0.000	0.006
<i>School</i>					
1. My child easily coped with schoolwork	8.65 (0.003)	0.009	12.21 (<0.001)	0.014	0.027

Table 1 continued

Subscales	Non-uniform DIF		Uniform DIF		
	χ^2 (P ^s)	ΔR^{2b}	χ^2 (P)	ΔR^2	$\frac{\beta_1^* - \beta_1}{\beta_1}$
2. My child enjoyed the school lessons	7.62 (0.005)	0.008	0.727 (0.39)	0.001	0.004
3. My child worried about his/her future	11.68 (<0.001)	0.012	4.38 (0.036)	0.005	0.002
4. My child was afraid of bad marks or grades	4.63 (0.031)	0.003	10.41 (0.001)	0.008	0.014

^a χ^2 is the value of the difference in $-2 \log$ likelihood of the models 1 and 2, and models 2 and 3 for testing uniform and non-uniform DIF, respectively

^b ΔR^2 is the R-square difference between the models 1 and 2, and models 2 and 3 for testing uniform and non-uniform DIF, respectively. S : p value. Bolded numbers are practically or statistically important

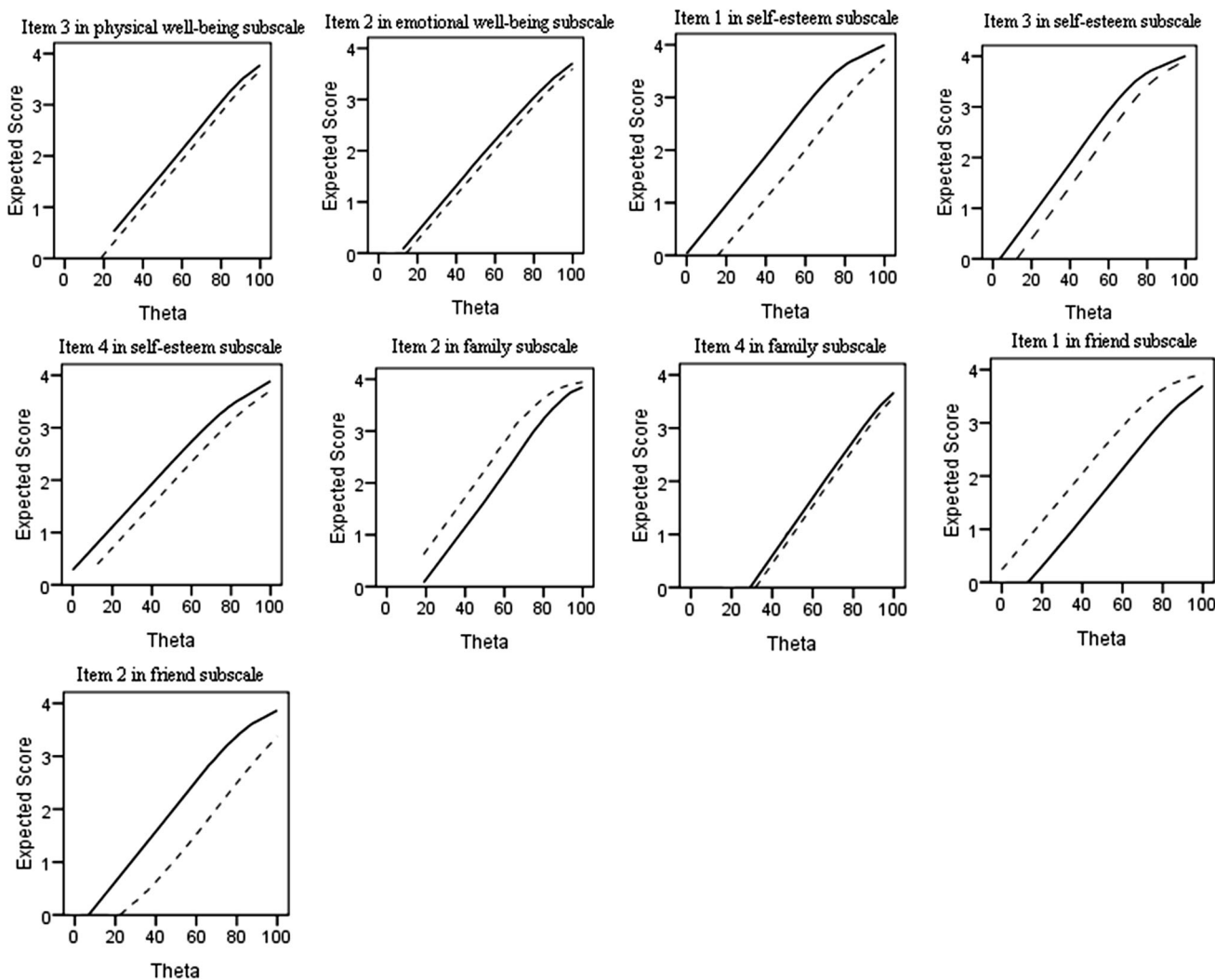


Fig. 1 Expected item score function of uniform DIF for Iranian (solid line) and Serbian (dashed line) children aged 8–12 (Kid-KINDL)

counterparts in all domains, except for the physical and friend subscales in the child self-reports for 8–12 year-old children, as well as in the school subscale for the parent proxy-reports for 8–12 year-old children. In order to assess

the potential impact of certain items with DIF in comparing HRQOL scores across two countries, items with uniform DIF, which were practically important, were removed from the subscales of the KINDL. Our finding revealed that

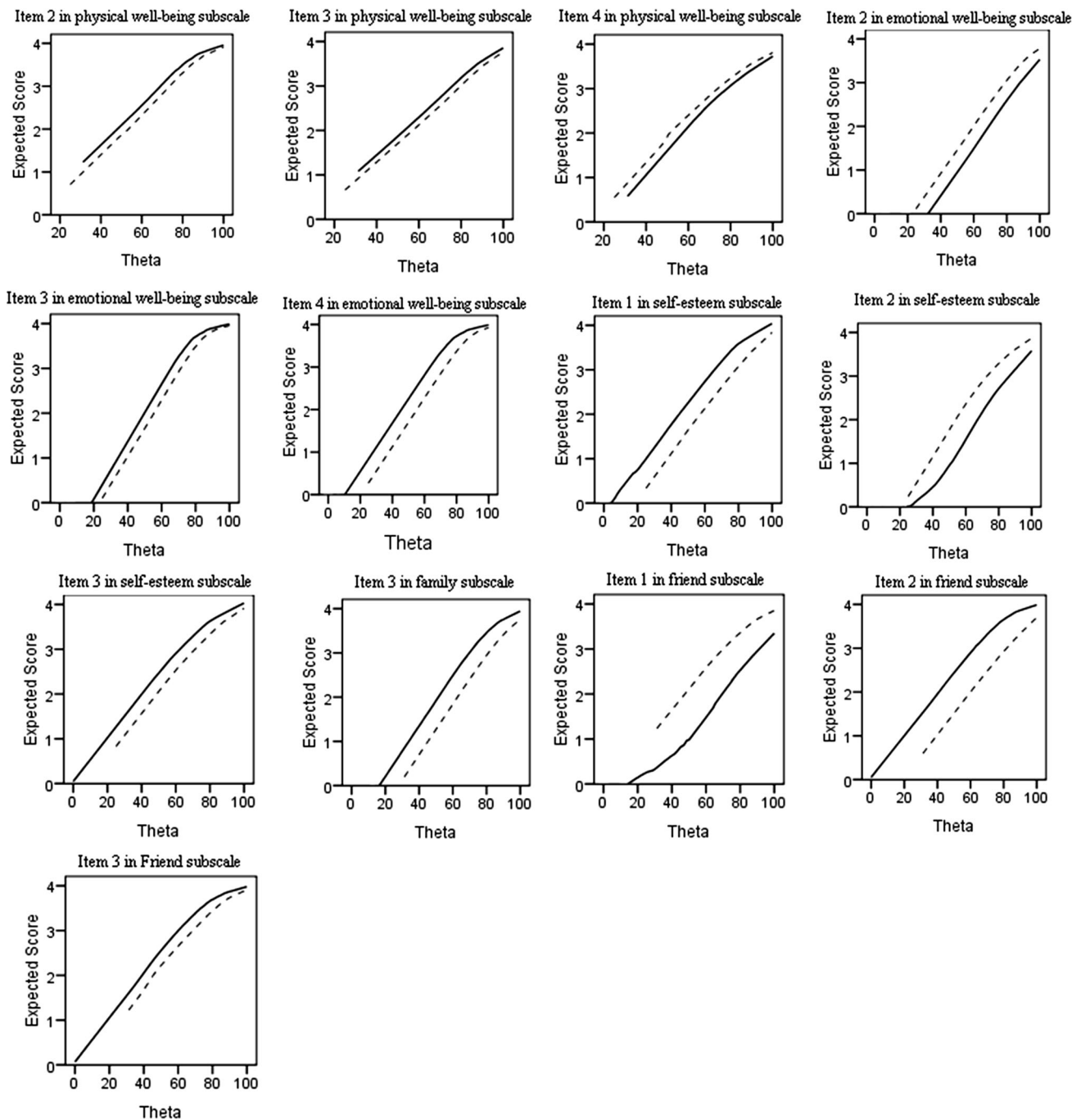


Fig. 2 Expected item score function of uniform DIF for Iranian (*solid line*) and Serbian (*dashed line*) parents aged 8–12 (Kid-KINDL)

ignoring or accounting for DIF items in the subscales had no considerable effects on group differences, except for the family subscale in child self-report of the Kid-KINDL. After removing item 1 “I felt fine at home” from family subscale in child self-report of the Kid-KINDL, mean scores across two countries were not statistically significant anymore.

Discussion

The importance of cross-cultural QOL comparisons is that they provide insight into the effects of cultural disparity pertaining to QOL scores across different countries or languages [30, 31]. However, a prerequisite for valid cross-cultural QOL comparisons is a questionnaire that

Table 2 The results of the ordinal logistic regression DIF analysis on the Kiddo-KINDL across Iranian and Serbian children (13–18)

Subscales	Non-uniform DIF		Uniform DIF		
	χ^2 ^a (P ^S)	ΔR^2 ^b	χ^2 (P)	ΔR^2	$\frac{\beta_1^* - \beta_1}{\beta_1}$
<i>Child self-reports</i>					
<i>Physical well-being</i>					
1. I felt ill	0.35 (0.55)	0.000	33.18 (<0.001)	0.017	0.003
2. I was in pain	0.56 (0.45)	0.000	16.99 (<0.001)	0.008	0.032
3. I was tired and worn-out	0.0004 (0.98)	0.000	31.76 (<0.001)	0.016	0.047
<i>Emotional well-being</i>					
1. I had fun and laughed a lot	0.075 (0.78)	0.000	21.78 (<0.001)	0.012	0.052
3. I felt alone	3.12 (0.07)	0.001	6.47 (0.009)	0.003	0.018
4. I felt scared or unsure of myself	16.39 (<0.001)	0.012	13.67 (<0.001)	0.009	0.076
<i>Self-esteem</i>					
1. I was proud of myself	2.81 (0.09)	0.001	146.72 (<0.001)	0.051	0.211
2. I felt on top of the world	0.318 (0.57)	0.000	188.16 (<0.001)	0.079	0.041
3. I felt pleased with myself	0.18 (0.66)	0.000	18.05 (<0.001)	0.007	0.052
4. I had lots of good ideas	0.56 (0.45)	0.000	35.47 (<0.001)	0.017	0.097
<i>Family</i>					
2. I felt fine at home	3.21 (0.07)	0.001	58.9 (<0.001)	0.024	0.019
3. We quarreled at home	1.72 (0.18)	0.001	20.26 (<0.001)	0.011	0.054
4. I felt restricted by my parents	4.88 (0.03)	0.003	36.33 (<0.001)	0.019	0.092
<i>Friends</i>					
1. I did things together with my friends	3.63 (0.06)	0.002	22.32 (<0.001)	0.011	0.054
2. I was a “success” with my friends	9.06 (0.002)	0.005	43.72 (<0.001)	0.023	0.114
4. I felt different from other children	20.38 (<0.001)	0.014	0.032 (0.85)	0.000	0.004
<i>School</i>					
1. Doing the schoolwork was easy	15.82 (<0.001)	0.009	0.34 (0.56)	0.000	0.002
2. I found school interesting	11.84 (<0.001)	0.007	8.81 (0.002)	0.006	0.012
3. I worried about my future	3.47 (0.062)	0.002	33.11 (<0.001)	0.016	0.014
4. I worried about getting bad marks or grades	9.93 (0.001)	0.004	17.08 (<0.001)	0.007	0.013
<i>Parent proxy-reports</i>					
<i>Physical well-being</i>					
3. My child was tired and worn-out	3.14 (0.07)	0.002	20.16 (<0.001)	0.010	0.052
4. My child felt strong and full of energy	0.54 (0.45)	0.000	19.75 (<0.001)	0.012	0.032
<i>Emotional well-being</i>					
1. My child had fun and laughed a lot	0.25 (0.62)	0.000	16.35 (<0.001)	0.009	0.046
2. My child didn't feel much like doing anything	1.36 (0.24)	0.001	10.61 (0.001)	0.006	0.035
4. My child felt scared or unsure of him-/herself	12.04 (<0.001)	0.007	78.89 (<0.001)	0.052	0.217
<i>Self-esteem</i>					
1. My child was proud of him-/herself	0.417 (0.52)	0.000	73.96 (<0.001)	0.028	0.133
2. My child felt on top of the world	1.56 (0.21)	0.001	90.12 (<0.001)	0.042	0.065
3. My child felt pleased with him-/herself	2.15 (0.14)	0.001	30.28 (<0.001)	0.012	0.077
4. My child had lots of good ideas	2.50 (0.11)	0.002	18.42 (<0.001)	0.011	0.075
<i>Family</i>					
2. My child felt fine at home	5.72 (0.02)	0.003	24.85 (<0.001)	0.012	0.034
3. We quarreled at home	3.98 (0.045)	0.002	97.79 (0.001)	0.056	0.214
4. My child felt that I was bossing him/her around	20.4 (<0.001)	0.011	6.23 (0.012)	0.004	0.0295
<i>Friends</i>					
1. My child did things together with friends	0.01 (0.91)	0.000	175.4 (<0.001)	0.075	0.084
2. My child was liked by other kids	2.41 (0.12)	0.001	239.08 (<0.001)	0.155	0.545

Table 2 continued

Subscales	Non-uniform DIF		Uniform DIF		
	χ^2 ^a (P ^s)	ΔR^2 ^b	χ^2 (P)	ΔR^2	$\frac{\beta_1^* - \beta_1}{\beta_1}$
3. My child got along well with his friends	0.12 (0.72)	0.000	26.50 (<0.001)	0.017	0.106
4. My child felt different from other children	18.59 (<0.001)	0.011	1.30 (0.25)	0.001	0.021
<i>School</i>					
1. My child easily coped with schoolwork	17.56 (<0.001)	0.012	13.76 (<0.001)	0.009	0.049
2. My child enjoyed the school lessons	4.53 (0.033)	0.003	7.56 (0.006)	0.005	0.033
3. My child worried about his/her future	10.12 (0.001)	0.004	41.03 (<0.001)	0.021	0.022
4. My child was afraid of bad marks or grades	1.47 (0.22)	0.001	15.77 (<0.001)	0.007	0.038

^a χ^2 is the value of the difference in -2 log likelihood of the models 1 and 2, and models 2 and 3 for testing uniform and non-uniform DIF, respectively

^b ΔR^2 is the R-square difference between the models 1 and 2, and models 2 and 3 for testing uniform and non-uniform DIF, respectively. \$: *p* value. Bolded numbers are practically or statistically important

demonstrates measurement invariance, so that subjects across different cultures perceive and respond to the questionnaire's items in the same or almost the same way [5].

The results of the present cross-cultural research showed that Iranian and Serbian children, as well as their parents, responded differently to the KINDL items. According to SR criterion, the majority of items in both the self- and proxy-reports of the Kid-KINDL and Kiddo-KINDL versions exhibited DIF. However, applying the effect size measures such as ΔR^2 and $\Delta\beta_1$ substantially decreased the detection rate of DIF. In addition, our findings revealed that the effect of DIF could not be cancelled out at the scale level for some subscales of the self- and proxy-reports. This finding implies that the direction of DIF for some items consistently favoured one group over another, resulting in significant scale-level bias. It was observed that, in the self-report of the Kid-KINDL, the effect of DIF could not be cancelled out for the physical, emotional and self-esteem subscales, and for the physical, emotional, self-esteem, friend and school subscales in the self-report of the Kiddo-KINDL. Moreover, in the parent proxy-reports, cancellation effect could not be obtained for the physical, emotional, self-esteem, family, and friend subscales of the Kid-KINDL, and for the emotional, self-esteem, friend, and school subscales in the Kiddo-KINDL. Nevertheless, it should be noted that full DIF cancellation at the subscale level rarely occurs in practice and should be interpreted with reserve [32]. Therefore, we should be cautious about comparing HRQOL subscale scores across Iranian and Serbian samples that have fulfilled the DIF cancellation criteria. Considered together, these findings indicate that the KINDL self- and proxy-reports do not produce invariant QOL measurements across Iranian and Serbian samples and cannot be used for cross-cultural comparisons, at least according to the findings of the present study. In this light, although Serbian

children and their parents rated children's QOL better than their Iranian counterparts across the domains, it is likely that the differences observed in QOL scores between the two countries reflected bias result [33]. It should be noted that techniques used for examination of the impact of DIF vary across DIF detection methods. In the present study, a removing and retaining strategy is used to determine how much mean group differences in subscale scores change with and without inclusion of items with DIF. In QOL instruments such as KINDL, where scales are often short, removing a number of items with DIF from the subscales may be a serious threat to the validity of the estimated subscale scores [28]. In response to this concern, we removed items with uniform DIF which were practically important. According to this strategy, our findings did not change in a meaningful way by ignoring or accounting for items with DIF.

An alternative way to test cross-cultural equivalence of the KINDL across two countries is structural equation modeling (SEM). A special case of SEM for detecting DIF is multiple-group confirmatory factor analysis (MGCFA). According to MGCFA, various types of measurement invariance hypotheses including configural, metric, and scalar equivalence can be tested. Configural invariance investigates whether an underlying construct of interest (in the present study six factors) is measured by the same set of items across two groups [34]. If configural test fails by showing poor fit indices, measurement invariance hypothesis is rejected and consequently latent means cannot be compared across groups [34].

Although there are no clear data about which factors lead to different interpretations of QOL items, possibilities include specific factors, such as values and perceptions about health. Different perceptions of well-being and functioning characterize QOL items across cultures. For example, the KINDL physical well-being domain generally

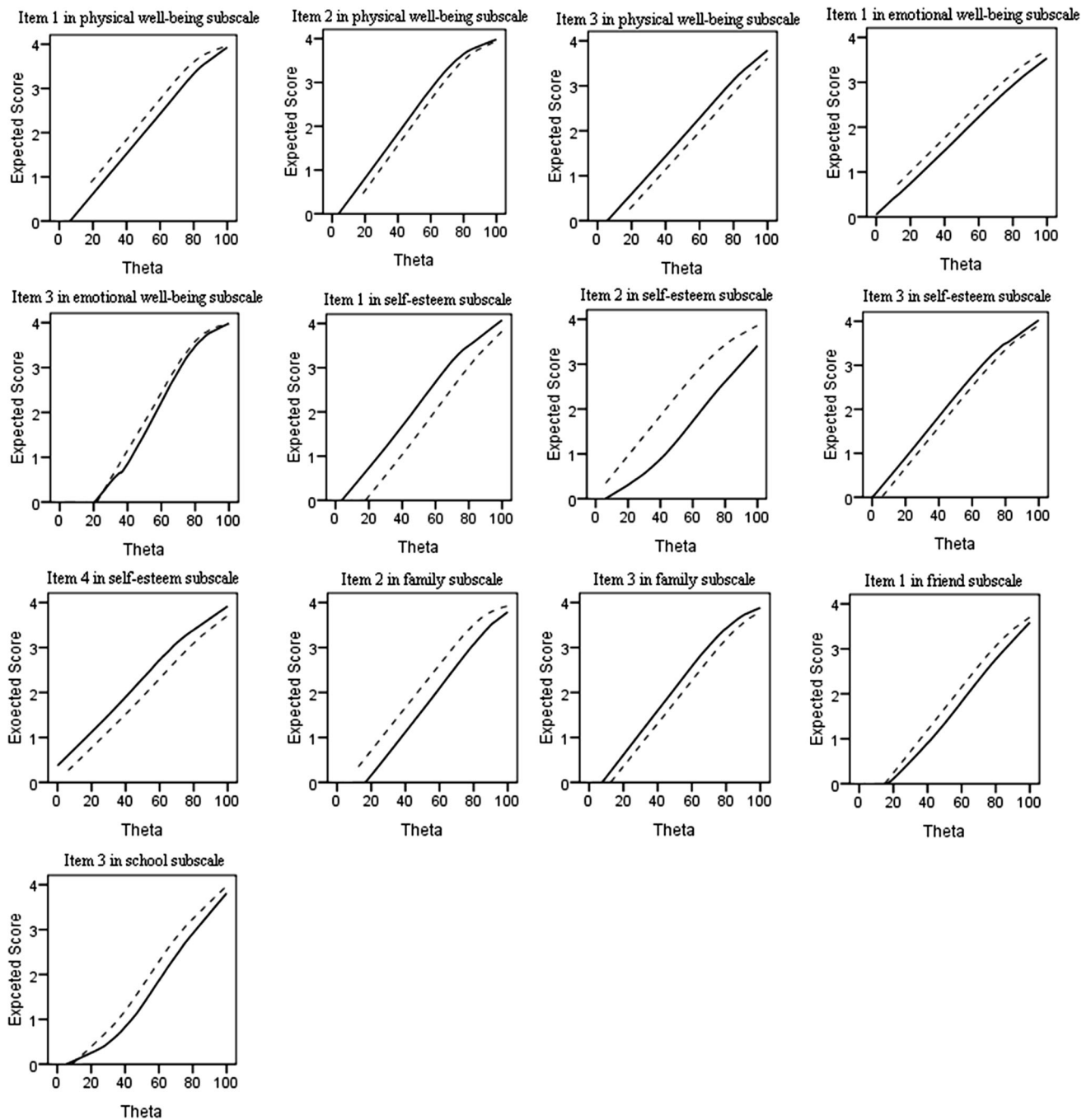


Fig. 3 Expected item score function of uniform DIF for Iranian (*solid line*) and Serbian (*dashed line*) children aged 13–18 (Kiddo-KINDL)

represents how one feels about physical abilities. It could be perceived as identifying whether one can physically function when not directly asked to do so, which signifies the physical functioning domain. Different values may be placed on QOL items across cultures. The KINDL family domain measures how a child functions within the family and is probably valued differently by a child from a Persian family that is more cohesive with more family members than, for example, by a child from a Western family with a

single parent. More general factors may include intrapersonal characteristics (e.g., ability to cope or independence), interpersonal characteristics (e.g., social relatedness or social support), or extra-personal characteristics (e.g., living conditions or financial resources). All of these, and many other factors, deserve further exploration by studies addressing cross-cultural measurement equivalence.

Since this is the first cross-cultural study organized with the aim of assessing the measurement equivalence of the

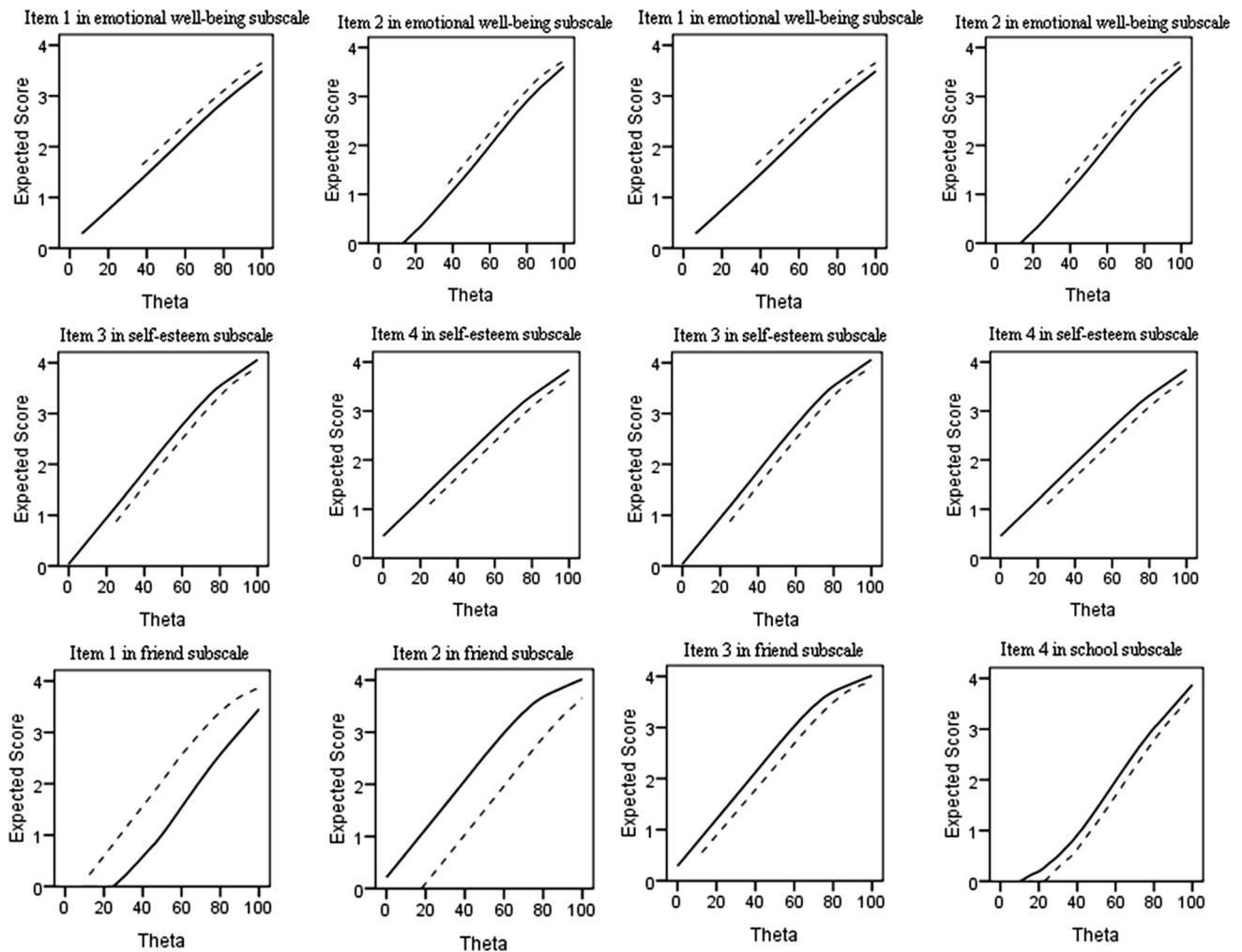


Fig. 4 Expected item score function of uniform DIF for Iranian (*solid line*) and Serbian (*dashed line*) parents aged 13–18 (Kiddo-KINDL)

KINDL self- and proxy-reports across two countries, there was no similar study in the literature for comparison. However, our findings were different from those in two previous studies, which reported substantial evidence for the cross-cultural equivalence of the KIDSCREEN-27 across 13 European countries [14, 15] and for the PedsQL™ 4.0 across four racial/ethnic subgroups in the United States [11]. Possible explanations for the differences between our findings and those of the two previous studies was likely due to the preconception that the KIDSCREEN had originally been developed using various cultures at the initial phases of the questionnaire's development, and due to the different statistical methods employed for invariance testing, or considering that only original language versions had been employed. For example, in the US study, the four ethnicity groups were compared within the same country and only children who had completed the PedsQL™ 4.0 in English were included in the study. Moreover, our findings were different from

those of previous research [15], which revealed that the KIDSCREEN-27 was equivalent across Iranian and Serbian children, alongside their parents. However, the selected samples in the present research are different from those of the KIDSCREEN study. This issue indicates that detecting DIF could vary substantially across different samples and from one measure to another.

One of the strengths of the present study is that the results of the DIF analysis were not affected by the child's age within and between cultures. In the present research, the Kid-KINDL (for 8–12 year-olds) and the Kiddo-KINDL (for 13–16 year-olds) were not combined to obtain a larger sample size for DIF analysis. If the two forms of the KINDL were merged while they had not been equivalent within one or both culture/language groups, it might have distorted the interpretation of the observed DIF at the national/language group level. Moreover, the effect of child's gender was controlled by taking it into account as a covariate in the OLR model, and the child's age was

Table 3 Comparison of QOL subscale scores of the KINDL across Iranian and Serbian samples

	Iranian sample Mean (SD)	Serbian sample Mean (SD)	T-value (<i>p</i> value)
<i>Self-report (8–12 years)</i>			
Physical well-being	79.27 (16.73)	78.51 (15.95)	0.60 (0.53)
Emotional well-being	81.23 (17.17)	84.07 (15.18)	2.33 (0.02)
Self-esteem	61.87 (22.75)	73.57 (17.49)	7.67 (<0.001)
Family	83.71 (15.18)	86.37 (12.98)	2.50 (0.012)
Friend	76.01 (18.46)	77.75 (15.78)	1.35 (0.17)
School	70.63 (22.74)	65.60 (19.83)	3.13 (0.002)
<i>Self-report (13–16 years)</i>			
Physical well-being	69.20 (19.44)	74.51 (16.67)	4.61 (<0.001)
Emotional well-being	70.41 (19.92)	83.79 (15.27)	11.68 (<0.001)
Self-esteem	55.11 (23.51)	70.81 (18.51)	11.55 (<0.001)
Family	76.89 (22.37)	87.06 (14.24)	8.22 (<0.001)
Friend	65.89 (18.28)	80.18 (16.18)	13.08 (<0.001)
School	51.46 (23.04)	54.09 (19.85)	1.99 (0.049)
<i>Proxy-report (8–12 years)</i>			
Physical well-being	77.32 (17.08)	81.38 (14.45)	3.17 (0.002)
Emotional well-being	75.57 (16.75)	86.21 (12.61)	8.79 (<0.001)
Self-esteem	66.57 (20.59)	75.39 (15.19)	5.98 (<0.001)
Family	76.73 (17.47)	86.28 (12.86)	7.63 (<0.001)
Friend	72.30 (17.03)	82.64 (12.90)	8.41 (<0.001)
School	69.58 (20.20)	72.42 (16.35)	1.91 (0.057)
<i>Proxy-report (13–16 years)</i>			
Physical well-being	69.07 (19.17)	78.79 (14.97)	8.19 (<0.001)
Emotional well-being	71.90 (18.54)	85.01 (13.79)	11.54 (<0.001)
Self-esteem	58.37 (22.05)	74.70 (14.40)	12.38 (<0.001)
Family	73.07 (21.13)	87.48 (12.98)	11.50 (<0.001)
Friend	66.11 (17.71)	82.79 (14.85)	14.98 (<0.001)
School	55.73 (21.46)	66.05 (17.54)	7.68 (<0.001)

automatically matched across two groups. Hence, child's age and gender were both balanced between Iranian and Serbian samples.

However, the present study had a number of potential limitations that should be taken into consideration when interpreting the results. First, because almost all children in Iranian and Serbian samples were healthy, the findings may not be generalized to children with chronic conditions. Second, differences in sample composition of the two countries with respect to important background variables such as parents' income, education, age, and gender can complicate the interpretation of observed DIF across two cultures. Because background variables (except gender and age) were not available in both samples, it was impossible to add them to the OLR model to explore whether they made a difference in DIF analysis results. This is especially important when we are trying to differentiate between true cultural disparities and differences in socio-demographic composition. Finally, it was not clear whether uncontrollable variables such as a parent and child's reading skills,

social desirability [35, 36] and acquiescence [37] contributed to the observed discrepancies across Iranian and Serbian samples.

Summary

The present study revealed that Iranian and Serbian children, as well as their parents, perceived and interpreted the meaning of almost all KINDL items differently. Although the directions of DIF showed that the effect of DIF could not be cancelled out at the scale level of some domains, further analysis revealed that removing or retaining the items with uniform DIF in subscales did not change our findings significantly when comparing HRQOL scores between the two countries. However, these findings indicate that self- and proxy-reports of the Kid-KINDL and Kiddo-KINDL have to be revised in order to be used for cross-cultural QOL comparisons. To gain a better insight into the measurement of non-invariance pertaining to the

KINDL for cross-cultural comparisons, further research is needed to test whether Iranian and Serbian versions of the KINDL are equivalent to the original version and other available versions.

References

- Acquadro C, Berzon R, Dubois D, Leidy NK, 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. *Value Health* 6:522–531
- Fayed N, de Camargo OK, Kerr E, Rosenbaum P, Dubey A, Bostan C et al (2012) Generic patient-reported outcomes in child health research: a review of conceptual content using World Health Organization definitions. *Dev Med Child Neurol* 54:1085–1095
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25:3186–3191
- Schmidt S, Bullinger M (2003) Current issues in cross-cultural quality of life instrument development. *Arch Phys Med Rehabil* 84:S29–S34
- Stewart AL, Napoles-Springer A (2000) Health-related quality-of-life assessments in diverse population groups in the United States. *Med Care* 38:II102–II124
- Sartorius N, Kuyken W (1994) Translation of health status instruments. In: Orley J, Kuyken W (eds) *Quality of life assessment: international perspectives*. Springer-Verlag, Berlin Heidelberg
- Regnault A, Herdman M (2015) Using quantitative methods within the Universalist model framework to explore the cross-cultural equivalence of patient-reported outcome instruments. *Qual Life Res* 24:115–124
- Ho SM, Rochelle TL, Law LS, Duan W, Bai Y, Shih SM et al (2014) Methodological issues in positive psychology research with diverse populations: exploring strengths among Chinese adults. In: Pedrotti JT, Edwards LM (eds) *Perspectives on the intersection of multiculturalism and positive psychology*. Springer Science & Business Media, Dordrecht
- Teresi JA, Fleishman JA (2007) Differential item functioning and health assessment. *Qual Life Res* 1:33–42
- Byrne BM, Watkins D (2003) The issue of measurement invariance revisited. *J Cross Cult Psychol* 34:155–175
- Limbers CA, Newman DA, Varni JW (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. *Ann Epidemiol* 19:575–581
- Ravens-Sieberer U, Auquier P, Erhart M, Gosch A, Rajmil L, Bruil J et al (2007) The KIDSCREEN-27 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. *Qual Life Res* 16:1347–1356
- Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Power M et al (2008) The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. *Value Health* 11:645–658
- Robitail S, Ravens-Sieberer U, Simeoni MC, Rajmil L, Bruil J, Power M et al (2007) Testing the structural and cross-cultural validity of the KIDSCREEN-27 quality of life questionnaire. *Qual Life Res* 16:1335–1345
- Stevanovic D, Jafari P (2015) A cross-cultural study to assess measurement invariance of the KIDSCREEN-27 questionnaire across Serbian and Iranian children and adolescents. *Qual Life Res* 24:223–230
- Varni JW, Burwinkle TM, Seid M, Skarr D (2003) The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambul Pediatr* 3:329–341
- 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. *Qual Life Res* 7:399–407
- Eser E, Yuksel H, Baydur H, Erhart M, Saatli G, Cengiz Ozyurt B et al (2008) The psychometric properties of the new Turkish generic health-related quality of life questionnaire for children (Kid-KINDL). *Turk Psikiyatri Derg* 19:409–417
- Fernandez-Lopez JA, Fernandez Fidalgo M, Cieza A, Ravens-Sieberer U (2004) Measuring health-related quality of life in children and adolescents: preliminary validation and reliability of the Spanish version of the KINDL questionnaire. *Aten Primaria* 33:434–442
- Helseth S, Lund T (2005) Assessing health-related quality of life in adolescents: some psychometric properties of the first Norwegian version of KINDL. *Scand J Caring Sci* 19:102–109
- Jafari P, Sharafi Z, Bagheri N, Shalileh S (2014) Measurement equivalence of the KINDL questionnaire across child self-reports and parent proxy-reports: a comparison between item response theory and ordinal logistic regression. *Child Psychiatry Hum Dev* 45:369–376
- Lee PH, Chang LI, Ravens-Sieberer U (2008) Psychometric evaluation of the Taiwanese version of the Kiddo-KINDL generic children's health-related quality of life instrument. *Qual Life Res* 17:603–611
- Stevanovic D (2009) Serbian KINDL questionnaire for quality of life assessments in healthy children and adolescents: reproducibility and construct validity. *Health Qual Life Outcomes* 7:79
- Stevanovic D, Lakic A, Vilotic J (2009) The psychometric study of the Serbian KINDL questionnaire for health-related quality of life assessment in children and adolescents. *Scand J Caring Sci* 23:361–368
- Yamaguchi N, Poudel KC, Poudel-Tandukar K, Shakya D, Ravens-Sieberer U, Jimba M (2010) Reliability and validity of a Nepalese version of the Kiddo-KINDL in adolescents. *Biosci Trends* 4:178–185
- Crane PK, Gibbons LE, Ocepek-Welikson K, Cook K, Cella D, Narasimhalu K et al (2007) A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Qual Life Res* 16(Suppl 1):69–84
- Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M (2010) Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health Qual Life Outcomes* 8:81
- Crane PK, Hart DL, Gibbons LE, Cook KF (2006) A 37-item shoulder functional status item pool had negligible differential item functioning. *J Clin Epidemiol* 59:478–484
- Huang IC, Leite WL, Shearer P, Seid M, Revicki DA, Shenkman EA (2011) Differential item functioning in quality of life measure between children with and without special health-care needs. *Value Health* 14:872–883
- Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 46:1417–1432

31. Schalock RL, Verdugo MA, Jenaro C, Wang M, Wehmeyer M, Jiancheng X et al (2005) Cross-cultural study of quality of life indicators. *Am J Ment Retard* 110:298–311
32. Wyse AE (2013) DIF cancellation in the Rasch Model. *J Appl Meas* 14:118–128
33. Kim SY, Nair R, Knight GP, Roosa MW, Updegraff KA (2008) Measurement equivalence of neighborhood quality measures for European American and Mexican American Families. *J Community Psychol* 37:1–20
34. Kim ES, Yoon M (2011) Testing measurement invariance: a comparison of multiple-group categorical CFA and IRT. *Struct Equ Modeling* 18:212–228
35. Lalwani AK, Shavitt S, Johnson T (2006) What is the relation between cultural orientation and socially desirable responding? *J Pers Soc Psychol* 90:165–178
36. Tourangeau R, Yan T (2007) Sensitive questions in surveys. *Psychol Bull* 133:859–883
37. Roessner V, Becker A, Rothenberger A, Rohde LA, Banaschewski T (2007) A cross-cultural comparison between samples of Brazilian and German children with ADHD/HD using the Child Behavior Checklist. *Eur Arch Psychiatry Clin Neurosci* 257:352–359