Original Research

Raters Interpret Positively and Negatively Worded Items Similarly in a Quality of Life Instrument for Children: Kid-KINDL

Chung-Ying Lin, PhD[1], Carol Strong, PhD[2],
Meng-Che Tsai, MD, MSc[2,3], and Chih-Ting Lee, MD[2,3]

Abstract

Measurement invariance is an important assumption to meaningfully compare children's quality of life (QoL) between different raters (eg, children and parents) and across genders. Moreover, QoL instruments may combine using negatively and positively worded items—a common method to reduce response bias. However, the wording effects may have different levels of impact on different raters and genders. Our aim was to investigate the measurement invariance of Kid-KINDL, a commonly used QoL instrument, across genders and raters and to consider the wording effects simultaneously. Third to sixth graders (208 boys and 235 girls) completed the self-rated Kid-KINDL, and 1 parent each of 241 children completed the parent-rated Kid-KINDL. The wording effects were accounted for by correlated traits-uncorrelated methods model. The measurement invariance was examined using multigroup confirmatory factor analysis. Item loadings and item intercepts were invariant across gender and rater when we simultaneously accounted for the wording effects of Kid-KINDL. Our results suggest that Kid-KINDL could be used to compare QoL across gender and that parent-rated Kid-KINDL could be used to measure children's QoL. Specifically, the invariant factor loadings across child-rated and parent-rated Kid-KINDL suggest that the score weights in each item were the same for both children and parents (ie, the important items identified by the children are the same items identified by the parents). The invariant item intercepts suggest that both children and parents share the same threshold for each item. Based on the results, we tentatively recommend that each score of a parent-rated Kid-KINDL can stand for each child's QoL.

Keywords

child, confirmatory factor analysis, factorial invariance, quality of life, wording effect

Introduction

Quality of life (QoL)—an important and unique patient self-reported assessment of one's health status—has been researched substantially and has been involved in clinical decision making.[1,2] Health care providers caring for pediatric populations also encourage the use of QoL in clinical evaluations; thus, a wide array of QoL procedures have been developed over the past few years.[3,4] Among the variety of QoL instruments for pediatric populations, KINDL is a multidimensional measurement of QoL, with physical, psychological, and social health subdomains. It is a feasible, reliable, and valid tool to measure an individual's QoL in line with the definition given by the World Health Organization.[5-9]

Several advantages of KINDL have been concluded by Lin et al.[10] The inclusion of both positively worded (eg, having fun and laughing a lot) and negatively worded (eg, feeling ill) items distinguishes KINDL from other pediatric QoL instruments. For example, Pediatric Quality of Life Inventory (PedsQL)—widely used in pediatric research—consists of negatively worded items only. In this case, participants may give biased reports because of response acquiescence and social desirability. Using both positively and negatively worded items in a single questionnaire can reduce ceiling effects, floor effects, and response bias.[11,12] Despite these potential benefits, the coexistence of positively and negatively worded items may threaten the internal consistency

[1]The Hong Kong Polytechnic University, Kowloon, Hong Kong
[2]National Cheng Kung University, Tainan, Taiwan
[3]National Cheng Kung University Hospital, Tainan, Taiwan

Corresponding Author:
Chung-Ying Lin, Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, 11 Yuk Choi Road, Hung Hom, Hong Kong.
Email: cylin36933@gmail.com

and validity of the instruments if wording effects are not considered.[13] However, the threat of wording effects on construct validity can be minimized and controlled by using a multitrait-multimethod (MTMM) approach combined with confirmatory factor analysis (CFA) models.[14]

The MTMM plus CFA approach compares several CFA models, where each model has its unique underlying construct.[14] In general, three types of models are compared, including positively worded and negatively worded items. We take the KINDL as an example as follows: (1) a conventional 6-QoL-factor model depending on the 5 dimensions from the KINDL, (2) a wording-method effect model, and (3) a model that takes into account both trait (dimensions in KINDL) and wording-method (positive and negative wording) factors.[10] Because the third type of model assists in separating method effects (an effect that is distinct from but may interfere with the trait effects) on the constructs of the instruments,[15-17] it can be used to verify construct validity of the instruments that have poor psychometric properties due to the wording effect. A previous study has shown that CFA models belonging to the third type[10]—especially the correlated traits-uncorrelated methods (CTUM) model (Figure 1)—have satisfactory fit indexes and outperform the other 2 types of models in the KINDL. Consequently, the KINDL has solid construct validity.

In addition to wording effects, factorial invariance (aka *measurement invariance*) is another important issue when developing and validating QoL instruments.[18,19] Before using an instrument made up with latent structures, a prerequisite is to examine the equivalence of the underlying constituent factors across different groups.[20,21] The cross-gender and rater (viz, parents and children) differences are particularly crucial for the use of pediatric QoL instruments.[22,23] Researchers need to ensure that boys and girls, and parents and children, have concordant interpretations of the pediatric QoL instrument.

Family or culture usually has different expectations for boys and girls;[24,25] thus, they may have different perspectives toward the same concepts. To prevent the discrepant interpretation of QoL constructs because of gender, we need to ensure that the cross-gender measurement invariance before comparison could be made between the QoL scores rated by different gender groups. Furthermore, health care providers sometimes need parents to rate their children's QoL when these children are too sick to report by themselves.[18] Health care providers may use the parent-rated scores for decision making. However, a wrong decision may be made if factorial invariance of a pediatric QoL instrument across parents and children does not exist.

Studies examining the factorial invariance of the KINDL questionnaire across different groups are sparse. To the best of our knowledge, only Jafari et al[26] have recently examined the cross-rater factorial invariance between parents and children. No studies have discussed the role of gender in the measurement invariance of the KINDL questionnaire. Furthermore, no studies have simultaneously examined the wording effects and

measurement invariance in the use of KINDL. Therefore, the purposes of the present study were as follows: (1) to examine the measurement invariance of Kid-KINDL across gender and (2) to examine the measurement invariance of Kid-KINDL across raters, simultaneously considering the wording effects.

Methods

Participants and Procedures

The Institutional Review Board of the National Cheng Kung University Hospital approved this study, and the data were collected from March to May 2010. Using a convenience sampling, the first author invited third- through sixth-grade students in 10 different schools (6 urban and 4 suburban) to participate in the study. All students completed a child-reported Kid-KINDL (version of KINDL questionnaires for children aged 8-12 years) in their classrooms under the supervision of their homeroom teachers and the first author. All parents filled out a parent-reported Kid-KINDL at home; the questionnaire was taken home by their children. The author collected the parent-reported Kid-KINDL from the students 1 to 3 days after each student filled out the child-reported Kid-KINDL.

Instruments. The Kid-KINDL questionnaire contains 6 dimensions (viz, physical well-being, emotional well-being, self-esteem, friends, family, and school), each of them including 4 items (Supplementary Table S1). The questionnaire was translated into Chinese for a pediatric population aged 8 to 12 years in Taiwan.[10] The Kid-KINDL can be obtained from the developers (www.kindl.org). Each item in the test is 5-point Likert-scaled (positively worded items: *never* scores 1 and *always* scores 5; negatively worded items: *never* scores 5 and *always* scores 1).[7] From the item scores, the total Kid-KINDL score and each dimension score can be calculated and transformed into a 0 to 100 result where a higher score represents better QoL.[7] Parallel child-reported and parent-reported Kid-KINDLs have been developed. Both questionnaires show acceptable psychometric properties,[5,27] including the Chinese version of Kid-KINDL for Taiwanese children.[10] Although studies found that some dimensions, including simultaneously both positively and negatively worded items, demonstrate unsatisfactory internal consistency,[7,8,28-30] some researchers argue that the unsatisfactory internal consistency does not necessarily threaten the construct validity of the Kid-KINDL.[10] However, it has been suggested that one item (Fr4: *felt different from other children*) be deleted from the Taiwanese version of Kid-KINDL due to its different meaning in Taiwanese and German, in which the original Kid-KINDL was developed.[10]

Data Analysis

Multigroup CFA was conducted to test the measurement invariance. CTUM models were used to examine invariance
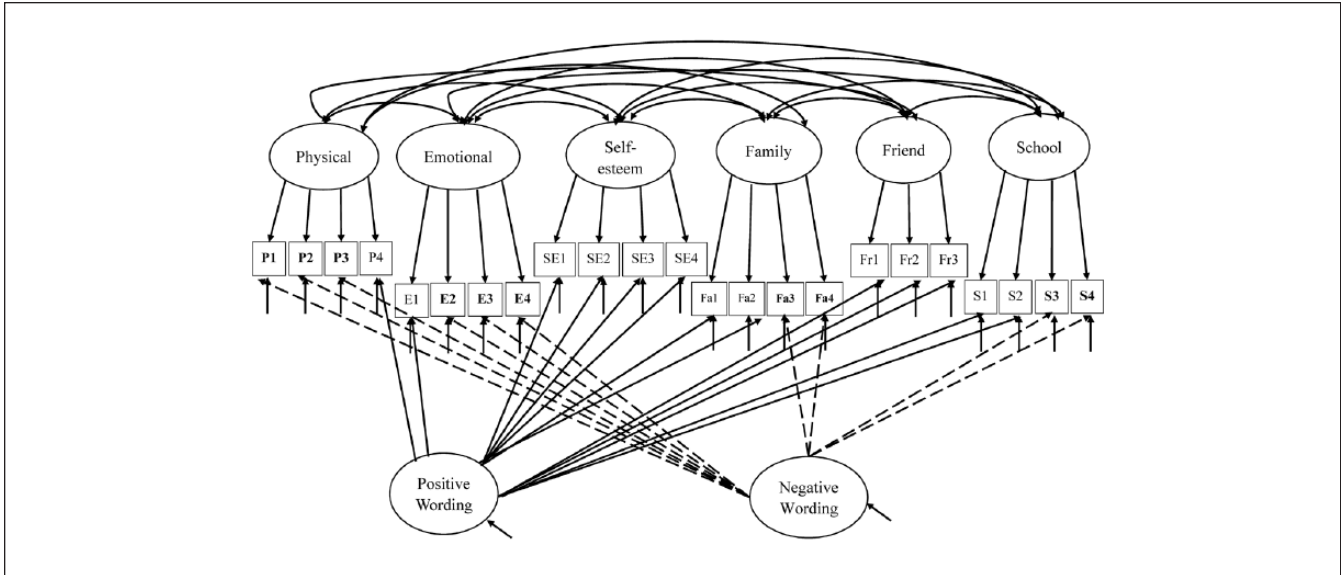
Figure 1. Correlated traits-uncorrelated methods framework represents 6 quality of life factors (physical, emotional, self-esteem, family, friends, and school) and 2 orthogonal wording factors (positive and negative wording).
Note. Negatively worded items are in bold font, and linked with Negative Wording concept by dashed lines.

across genders (boys and girls) and raters (parents and children). Because fathers and mothers may also have different interpretations toward the same item description, we further performed sensitivity analyses to test the measurement invariance across mothers and children. We used only mothers for the examination because parent-proxies are often completed by mothers; the number of fathers was not sufficient for conducting CFA (n = 55). Four baseline CTUM models were conducted separately for boys, girls, parents, and children to examine the goodness of fit of the models to the data for each group. The CTUM models separate the underlying concepts of each item into two parts: trait and method. The traits in the CTUM models are the QoL concepts in the Kid-KINDL (physical well-being, emotional well-being, self-esteem, friends, family, and school). The methods in the models are the wording effects (the positively and negatively worded items). The 6 QoL concepts were correlated in the model; the methods were not correlated (Figure 1). Moreover, the QoL concepts and the method effects were not correlated in the model because we assumed that QoL concepts and method effects are different and uncorrelated underlying constructs.

The following fit indexes were used to evaluate the fit of our CTUM models: the acceptable fit indexes according to published guidelines, a goodness of fit index (GFI), a non-normed fit index (NNFI) and a comparative fit index (CFI) > 0.9. A root mean square error of approximation (RMSEA) and a standardized root mean square residual (SRMR) < 0.08 indicate acceptable model fit.[31,32]

We considered 4 kinds of measurement invariance in our analyses: configural, metric invariance, scalar, and uniqueness.[21] Configural invariance indicates conceptual

equivalence of the underlying theoretical factors across groups, metric invariance suggests equivalent associations between items and latent factors across groups, scalar invariance means the same origins across groups (the groups have their scores rated unbiased), and uniqueness invariance suggests that items are influenced to the same degree and perhaps by the same unique factors across groups.[21]

Based on the suggestions mentioned above, 3 configural CTUM models (1 for gender and 2 for rater groups) were performed to test the equivalence of the overall structure (6 correlated QoL traits and 2 uncorrelated wording effects). In addition, 3 nested models were compared with each other; each model includes the restrictions from the previous models with the additional constraints of equal factor loadings, equal item intercepts, and equal item uniqueness across gender and rater groups. When comparing these nested models, Δ values >−0.01 for CFI and <0.015 for the RMSEA indicate that the null hypothesis of invariance should not be rejected.[33] Although some researchers[19,34-36] state an alternative Δ value >−0.02 for CFI, we intended to use ΔCFI values >−0.01 for measurement invariance.

LISREL 8.8 (Scientific Software International Inc, Lincolnwood, Illinois) was used to perform CFAs; other analyses were done using SPSS 16.0 (SPSS Inc, Chicago, Illinois).

Results

The mean age of the children (n = 443) was 10.57 (SD = 1.08) years; 47.0% of them were boys. The mean ages of fathers and mothers were 43.62 (SD = 5.70) and 40.81 (SD = 5.18) years, respectively. Nearly 30% of the fathers had

Table 1. Demographics of Children and Their Parents.

	Children (n = 443)	Father (n = 185)[a]	Mother (n = 183)[a]
Age, mean ± SD, y	10.57 ± 1.08	43.62 ± 5.70	40.81 ± 5.18
Gender, n (%), male	208 (47.0)	173 (100.0)	0 (0.0)
Grade, n (%)			
Third	97 (21.9)	—	—
Fourth	116 (26.2)	—	—
Fifth	150 (33.9)	—	—
Sixth	80 (18.1)	—	—
Education, n (%)			
Elementary	—	7 (2.9)	1 (0.4)
Junior high	—	12 (5.0)	6 (2.5)
Senior high	—	71 (29.5)	92 (38.2)
College	—	69 (28.7)	78 (32.3)
Graduate school	—	26 (10.8)	6 (2.5)

[a]With missing values on demographic data.

senior high education (29.5%) and college (28.7%), while more than 30% of the mothers had senior high education (38.2%) and college (32.3%) (Table 1). Also, as we anticipated, most parent-proxies were completed by mothers (n = 176, 73.0%).

Except for 1 nonsignificant factor loading (item S3 in parent groups; $P = 0.19$), all other item factor loadings were significant in the baseline CTUM models among all groups: boy (n = 208), girl (n = 235), parent (n = 241), mother (n = 176), and children (n = 443) (Supplementary Table S2). Furthermore, the baseline CTUM models show satisfactory or approximately acceptable data-model fit for all groups, including global configural models in gender (GFI = 0.894 to 0.900, NNFI = 0.948 to 0.965, CFI = 0.960 to 0.974, RMSEA = 0.0476 to 0.0491, SRMR = 0.0547 to 0.0609) and rater groups (GFI = 0.882 to 0.934, NNFI = 0.965 to 0.970, CFI = 0.974 to 0.977, RMSEA = 0.0438 to 0.0478, SRMR = 0.0452 to 0.0535), except for the significant χ^2 values (Supplementary Table S3).

Based on the good fit indexes in each CTUM model, the following measurement invariance tests were affirmative. Measurement invariance was supported by item loadings, item intercepts, and item uniqueness across gender (ΔCFI = −0.007 to −0.001 and ΔRMSEA = −0.0004 to 0.0029). Item loadings and item intercepts were invariant across raters (ΔCFI = −0.014 to −0.009 and ΔRMSEA = 0.0057 to 0.0079 using all parents' parent-proxies; ΔCFI = −0.014 to −0.008 and ΔRMSEA = 0.0058 to 0.0062 using mothers' parent-proxies only), but item uniqueness was not (ΔCFI = −0.062 and ΔRMSEA = 0.0199 using all parents' parent-proxies; ΔCFI = −0.045 and ΔRMSEA = 0.0263 using mothers' parent-proxies only) (Table 2).

Discussion

This study examines the measurement invariance of the Chinese version of the Kid-KINDL across genders and raters

among Taiwanese children. The most important finding was that the underlying concepts of Kid-KINDL (QoL-trait and wording-method effects)[7,10] are equivalent across boys and girls, and across parents and children. Although we found that item uniqueness was not invariant across raters, this finding does not contradict our conclusion that underlying concepts of Kid-KINDL are equivalent across parents and children. Item uniqueness indicates the measurement errors of each item, which is not essential for measurement invariance.

Our results show that boys and girls in Taiwan have similar interpretations of the items on the Kid-KINDL[7,10] under the same impact of wording effects. This indicates that they can capture the concepts of Kid-KINDL in a very similar way, thus making their scores comparable across genders, instead of viewing them separately. Studies have often used Kid-KINDL to measure and analyze children's QoL without distinguishing between boys and girls.[5,28-30] Our current results support that the merged use is reasonable. Although our results seem to be the first to detect the measurement invariance across genders for Kid-KINDL, our results could somewhat correspond to other studies which have found measurement invariance in different QoL instruments, such as PedsQL.[18] Our results also show satisfactory data-model fit in the CTUM model, which agrees with the statement that the wording effects exist in Kid-KINDL.[10,37]

Several researchers state that standardized QoL instruments should explore whether the test items are interpreted similarly across age, gender, language, socioeconomic, and race/ethnicity subpopulations.[20,36,38] Confirmation of strict levels of factorial invariance, as provided by the present study, is especially suggested. That is why factorial invariance studies should be repeated periodically due to the change of subpopulations over time.[20] Due to the characteristics of our sample—sample size was not large enough for participants with different ages, our data did not include socioeconomic information, all participants used the same language and were the same race/ethnicity—we could only

Table 2. Measurement Invariance Across Gender and Raters.

	χ^2	$\Delta\chi^2$	df	Δdf	CFI	ΔCFI	ΔRMSEA	RMSEA
Gender								
Model 1G	622.433*	—	384	—	0.969	—	0.0484	—
Model 2G − 1G	—	111.15*	—	61	—	−0.007	—	0.0022
Model 3G − 2G	—	21.27	—	15	—	−0.001	—	−0.0004
Model 4G − 3G	—	58.88*	—	23	—	−0.004	—	0.0029
Rater								
Model 1R	663.802*	—	384	—	0.976	—	0.0444	—
Model 2R − 1R	—	224.80*	—	61	—	−0.014	—	0.0079
Model 3R − 2R	—	122.89*	—	15	—	−0.009	—	0.0057
Model 4R − 3R	—	746.11*	—	23	—	−0.062	—	0.0199
Rater[a]								
Model 1M	653.840*	—	384	—	0.975	—	0.0456	—
Model 2M − 1M	—	209.68*	—	61	—	−0.014	—	0.0062
Model 3M − 2M	—	109.14*	—	15	—	−0.008	—	0.0058
Model 4M − 3M	—	129.76*	—	23	—	−0.045	—	0.0263

Note. Models 1G, 1R, and 1M: Configural model. Models 2G, 2R, and 2M: Constrained all item loadings equal across gender (Model 2G) and raters (Models 2R and 2M). Models 3G, 3R, and 3M: Constrained all item loadings and item intercepts equal across gender (Model 3G) and raters (Models 3R and 3M). Models 4G, 4R, and 4M: Constrained all item loadings, item intercepts, and item uniqueness equal across gender (Model 4G) and raters (Models 4R and 4M). Models 2G, 3G, and 4G were based on Model 1G with additional constraints; Models 2R, 3R, and 4R were based on Model 1R with additional constraints; Model 2M, 3M, and 4M were based on Model 1M with additional constraints. CFI = comparative fit index; RMSEA = root mean square error of approximation.
[a]Parent-proxies only used mothers' reports.
*p < .05.

test the measurement invariance across gender. Future studies on other factors are needed.

However, our study provided information on the factorial invariance of another important factor: the rater. Some evidence shows that children aged 3 to 8 years can understand response terms and underlying concepts of QoL.[39] They are able to self-rate their QoL, but valuable supplementary information for health care providers can be gathered by parent reports on QoL. For example, a parent-rated QoL proxy can be used as a substitute for a child who is too young or too ill to complete a self-report.[34] In addition, health care providers can evaluate the agreement of children's QoL between parents' and children's perceptions and use this information on agreement to improve the children's QoL.[24,40-42] Our results support the assumption of measurement invariance across children and parents for the Kid-KINDL; therefore, the use of a parent-rated Kid-KINDL is justified. As a result, health care professionals may want to use Kid-KINDL for their clinical decision making.

Nonetheless, health care professionals should still use the parent-rated Kid-KINDL with caution because other studies[5,26] on KINDL demonstrated results that were inconsistent with ours. For example, Erhart et al[5] concluded that parent-rated KINDL achieved slightly higher reliability; Jafari et al[26] found that 12 KINDL items may be noninvariant across children and parents. Two reasons may explain the different results: First, Erhart et al[5] and Jafari et al[26] did not consider the wording effects for KINDL and may have their results somewhat different from ours; second, Jafari et al[26] used different methods from ours (item response theory

[IRT] and ordinal logistic regression [OLR] methods) to examine the factorial invariance; Erhart et al[5] did not conduct a factorial invariance analysis. Because different methods on factorial invariance may have different findings,[26,43] future studies examining factorial invariance with different factorial invariance methods on a large sample size are warranted. Specifically, IRT methods usually need a sample size over 500 to achieve the correct estimations.[44]

In addition, the 3 methods on factorial invariance (viz, IRT, OLR, and multigroup CFA) have their own advantages and disadvantages. For IRT and multigroup CFA, both share the advantage of nonuniform factorial invariance; that is, they estimate both item intercepts and factor loadings for invariance, while OLR cannot.[26,45] Both OLR and multigroup CFA have the advantage of adjusting confounders in the model; that is, they can allow additional variable in the model, while IRT cannot.[26] Multigroup CFA, however, treats the items as continuous variables, which is apparently not the case in the Kid-KINDL that adopts Likert-type scales.[26]

This study has some limitations. First, our sample may not be representative of the pediatric population in Taiwan because we did not recruit children with any chronic conditions. Therefore, our results may not be generalized beyond healthy children. Second, the measurement invariance was only examined for CTUM models, while there are different kinds of models that can serve to examine the construct validity that account for wording effects in addition to the CTUM model, such as the correlated trait-correlated methods model and correlated trait-correlated methods minus one model. We only demonstrated that boys and girls, and

parents and children, understand the items of the Kid-KINDL and understand the construct behind the Kid-KINDL in the same way under the CTUM model. We cannot guarantee that our results can be generalized to other models accounting for wording effects. However, the CTUM model was found to be the best model among all models accounting for wording effects.[10] We justified that only the test in the CTUM model is appropriate. Third, the examination of factorial invariance across parents and children did not consider the gender of the responding parent.

Although our results support the measurement invariance across boys and girls, the inference of gender invariance of the parents should be verified before applying to fathers' and mothers' report. Therefore, the merged use of fathers' and mothers' reports as a single proxy rating to represent the *parents'* rating may weaken the verification of measurement invariance across parents and children. Therefore, we examined the measurement invariance across mothers and children and showed that the measurement invariance was supported. However, sample size was not sufficient in our study to test the measurement invariance across fathers and children. Therefore, additional studies to examine the measurement invariance across fathers and children in Kid-KINDL are highly recommended.

Researchers may also want to test the measurement invariance across fathers and mothers. Following the third limitation, the small sample size in our study should be noted. The recommended sample size for CFA is 200 per group, which indicates that our mother-rated Kid-KINDLs were slightly fewer than the sufficient number. Finally, the Kid-KINDL is designed for children 8 to 12 years of age; findings for other age versions of KINDL questionnaires could differ from our results. Future research on other age versions of KINDL questionnaires (eg, Kiddo-KINDL for children 13 to 16 years of age) is thus required.

Conclusion

Despite the limitations, this study demonstrated that Kid-KINDL exhibits good psychometric properties, including wording effects and measurement invariance across gender groups and between parents and children. This study provided support for clinicians to use Kid-KINDL among boys and girls. More importantly, this study provides a foundation to use parent-reported QoL in necessary situations, when children are unable to report by themselves. Moreover, the tested Kid-KINDL is a 23-item Chinese version; researchers may want to know whether the original 24-item version has similar findings to ours.

Acknowledgments

Declaration of Conflicting Interests

Funding

Supplementary Material

The supplementary materials are available at http://journals.sage-pub.com/doi/suppl/10.1177/0046958017696724

References

1. European Medicines Agency. Committee for medicinal products for human use (CHMP). Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf. Published 2005. Accessed October 25, 2016.

2. Su C-T, Ng H-S, Yang A-L, Lin C-Y. 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. *Psychol Assess*. 2014;26(3):980-989.

3. Ravens-Sieberer U, Erhart M, Wille N, Wetzel R, Nickel J, Bullinger M. Generic health-related quality-of-life assessment in children and adolescents: methodological considerations. *Pharmacoeconomics*. 2006;24(12):1199-1220.

4. Varni JW, Burwinkle TM, Lane MM. Health-related quality of life measurement in pediatric clinical practice: an appraisal and precept for future research and application. *Health Qual Life Outcomes*. 2005;3:34.

5. Erhart M, Ellert U, Kurth BM, Ravens-Sieberer U. Measuring adolescents' HRQoL via self reports and parent proxy reports: an evaluation of the psychometric properties of both versions of the KINDL-R instrument. *Health Qual Life Outcomes*. 2009;7:77.

6. Helseth S, Lund T. Assessing health-related quality of life in adolescents: some psychometric properties of the first Norwegian version of KINDL. *Scand J Caring Sci*. 2005;19(2):102-109.

7. Ravens-Sieberer U, Bullinger M. KINDLR-questionnaire for measuring health-related quality of life in children and adolescents: revised version manual. http://www.kindl.org/daten/pdf/ManEnglish.pdf. Published 2000. Accessed October 25, 2016.

8. Stevanovic D. Serbian KINDL questionnaire for quality of life assessments in healthy children and adolescents: reproducibility and construct validity. *Health Qual Life Outcomes*. 2009;7:79.

9. World Health Organization. Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL). *Qual Life Res*. 1993;2(2):153-159.

10. Lin C-Y, Luh W-M, Cheng C-P, Yang A-L, Ma H-I. Evaluating the wording effect and psychometric properties of the Kid-KINDL: using the multitrait-multimethod approach. *Eur J Psychol Assess*. 2014;30(2):100-109.

11. Anastasi A. *Psychological Testing*. 5th ed. New York, NY: Macmillan; 1982.

12. Nunnally J. *Psychometric Theory*. 2nd ed. New York, NY: McGraw-Hill; 1978.

13. Podsakoff PM, Mackenzie SB, Lee J-Y, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol*. 2003;88(5):879-903.

14. Motl RW, Distefano C. Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Struct Equ Modeling*. 2002;9(4):562-578.

15. Eid M. A multitrait-multimethod model with minimal assumptions. *Psychometrika*. 2000;65(2):241-261.

16. Marsh HW, Grayson D. Latent variable models of multitrait-multimethod data. In: Hoyle RH, ed. *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage; 1995:177-198.

17. Marsh HW, Hocevar D. Confirmatory factor analysis of multitrait-multimethod matrices. *J Educ Meas*. 1983;20(3):231-248.

18. Lin C-Y, Luh W-M, Cheng C-P, Yang A-L, Su C-T, Ma H-I. 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 Hum Dev*. 2013;44(5):583-590.

19. Lin C-Y, Luh W-M, Yang A-L, Su C-T, Wang J-D, Ma H-I. 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. *Qual Life Res*. 2012;21(1):177-182.

20. Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? testing measurement invariance using the confirmatory factor analysis framework. *Med Care*. 2006;44(11)(suppl 3):S78-S94.

21. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ Res Methods*. 2000;3:4-70.

22. Bagheri Z, Jafari P, Tashakor E, Kouhpayeh A, Riazi H. Assessing whether measurement invariance of the KIDSCREEN-27 across child-parent dyad depends on the child gender: a multiple group confirmatory factor analysis. *Glob J Health Sci*. 2014;6(5):142-153.

23. Jafari P, Bagheri Z, Hashemi SZ, Shalileh K. Assessing whether parents and children perceive the meaning of the items in the PedsQL™ 4.0 quality of life instrument consistently: a differential item functioning analysis. *Glob J Health Sci*. 2013;5(5):80-88.

24. Su C-T, Wang J-D, Lin C-Y. Child-rated versus parent-rated quality of life of community-based obese children across gender and grade. *Health Qual Life Outcomes*. 2013;11:206.

25. Witt S. Parental influence on children's socialization to gender roles. *Adolescence*. 1997;32(126):253-259.

26. Jafari P, Sharafi Z, Bagheri Z, Shalileh S. 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*. 2014;45(3):369-376.

27. Ravens-Sieberer U, Bullinger M. Assessing health-related quality of life in chronically ill children with the German KINDL: first psychometric and content analytical results. *Qual Life Res*. 1998;7(5):399-407.

28. Lee P-H, Chang L-I, Ravens-Sieberer U. Psychometric evaluation of the Taiwanese version of the Kiddo-KINDL generic children's health-related quality of life instrument. *Qual Life Res*. 2008;17(4):603-611.

29. Wee HL, Lee WWR, Ravens-Sieberer U, Erhart M, Li SC. Validation of the English version of the KINDL generic children's health-related quality of life instrument for an Asian population—results from a pilot test. *Qual Life Res*. 2005;14(4):1193-1200.

30. Wee HL, Ravens-Sieberer U, Erhart M, Li SC. Factor structure of the Singapore English version of the KINDL children quality of life questionnaire. *Health Qual Life Outcomes*. 2007;5:4.

31. Hoyle RH, Panter AT. Writing about structural equation modeling. In: Hoyle RH, ed. *Structural Equation Modeling: Concepts, Issues, and Application*. Thousand Oaks, CA: Sage; 1995:158-176.

32. Kline RB. *Principles and Practice of Structural Equation Modeling*. 2nd ed. New York, NY: Guilford Press; 2005.

33. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Modeling*. 2007;14(3):464-504.

34. Limbers CA, Newman DA, Varni JW. 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 Health*. 2008;11(4):659-668.

35. Limbers CA, Newman DA, Varni JW. 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*. 2009;19(8):575-581.

36. Varni JW, Limbers CA, Newman DA. Factorial invariance of the PedsQL™ 4.0 Generic Core Scales self-report across gender: a multigroup confirmatory factor analysis with 11,356 children ages 5 to 18. *Appl Res Qual Life*. 2008;3(2):137-148.

37. Lee C-T, Lin C-Y, Tsai M-C, Strong C, Lin Y-C. Psychometric evaluation and wording effects on the Chinese version of the parent-proxy Kid-KINDL. *Health Qual Life Outcomes*. 2016;14:123.

38. Meredith AW, Teresi AJ. An essay on measurement and factorial invariance. *Med Care*. 2006;44(11)(suppl 3):S69-S77.

39. Cremeens J, Eiser C, Blades M. Characteristics of health-related self-report measures for children aged three to eight years: a review of the literature. *Qual Life Res*. 2006;15(4):739-754.

40. Eiser C, Morse R. Can parents rate their child's health-related quality of life? results of a systematic review. *Qual Life Res*. 2001;10(4):347-357.

41. Lin C-Y, Su C-T, Wang J-D, Ma H-I. Self-rated and parent-rated quality of life for community-based obese and overweight children. *Acta Paediatr*. 2013;102(3):e114-e119.

42. Upton P, Lawford J, Eiser C. Parent-child agreement across child health-related quality of life instruments: a review of the literature. *Qual Life Res*. 2008;17(6):895-913.

43. Yang FM, Heslin KC, Mehta KM, et al. A comparison of item response theory-based methods for examining differential item functioning in object naming test by language of assessment among older Latinos. *Psychol Test Assess Model*. 2011;53:440-460.

44. Chang C-C, Lin C-Y, Gronholm PC, Wu TH. Cross-validation of two commonly used self-stigma measures, Taiwan versions of the Internalized Stigma Mental Illness Scale and Self-Stigma Scale-Short, for people with mental illness [published online ahead of print July 5, 2016]. *Assessment*. doi:10.1177/1073191116658547.

45. Wiesner M, Windle M, Kanouse DE, Elliott MN, Schuster MA. DISC Predictive Scales (DPS): factor structure and uniform differential item functioning across gender and three racial/ethnic groups for ADHD, conduct disorder, and oppositional defiant disorder symptoms. *Psychol Assess*. 2015;27(4):1324-1336.