Psychometric properties of the Mandarin Chinese version of the KIDSCREEN-52 health-related quality of life questionnaire in adolescents: a cross-sectional study

Yuhang Zhu[1] · Juan Li[1] · Shanju Hu[1] · Xiaoxue Li[2] · Di Wu[3] · Shuai Teng[1]

Abstract

Purpose To evaluate the psychometric properties of the Mandarin Chinese version of KIDSCREEN-52 health-related quality of life (HRQOL) questionnaire.

Methods We conducted a cross-sectional study with two cohorts of school-based data in the urban areas of Weifang, China. 4385 and 841 representative adolescents aged 11–17 years have recruited into the baseline and test–retest group, respectively. Psychometric analyses included feasibility, item and dimension properties, reliability, construct validity, measurement invariance, convergent and discriminant validity, and known-group validity.

Results The response rates of both baseline and test–retest surveys were more than 90%. Low missing values were found (0.02–1.92%) across ten dimensions. The measurement properties of items were satisfactory. Noteworthy ceiling effects were observed for ten dimensions (6.75–31.84%), while the observed floor effects were negligible (0.02–1.37%). Internal consistency was robust with Cronbach's alpha (0.819–0.959), while the test–retest reliability was acceptable with the ICCs (0.724–0.849). Confirmatory factor analysis confirmed the ten-dimensional structure and supported the configural and metric invariance across gender and age groups. The Pearson correlation coefficients between the KIDSCREEN-52 and the PedsQL™ 4.0 were stronger in comparable dimensions than those in less comparable dimensions, demonstrating the convergent and discriminant validity. In most dimensions, statistically significant and medium or large effect size differences were found across socioeconomic and mental health status, supporting the known-group validity.

Conclusion These findings demonstrate that the Mandarin Chinese version of KIDSCREEN-52 seems to work well in a Chinese context, and is a psychometrically valid and reliable HRQOL instrument.

Keywords Adolescents · Health-related quality of life · Mandarin Chinese · Psychometric properties · KIDSCREEN-52

Introduction

Quality of life (QOL) is a complex concept. It has been defined by the WHO Quality of Life Group as an individual's perception of well-being (physical, psychological, and social) in the context of their values, beliefs, expectations, goals, and cultural environment [1]. Nowadays, the term "health-related quality of life (HRQOL)" is more common in the medical and health fields. As a subjective, multidimensional and dynamic health screening indicator [2], HRQOL plays an increasingly important role in clinical trials, epidemiological and health policy settings of children and adolescents [3–6]. In clinical trials, HRQOL measures can determine the burden of diseases or disabilities [7]. Furthermore, the combination of clinical data and HRQOL measures can provide a complete assessment of the evaluating health care

✉ Yuhang Zhu
  zhuyuhang369@163.com

1 School of Public Health and Management, Weifang Medical University, No. 7166 Baotong West Street, Weifang, Shandong, People's Republic of China

2 Institute of Modern Hospital Management, The Third Medical Center of Chinese PLA General Hospital, Beijing, People's Republic of China

3 Department of Teaching Quality Monitoring and Evaluation, Weifang Medical University, Weifang, Shandong, People's Republic of China

interventions and treatments for overall well-being of children and adolescents [8]. From an epidemiological point of view, HRQOL research conducted within a general population can help identify at-risk children and adolescents in schools and communities. These risks include in bullying, developmental disorders, poor peer relations, domestic violence, disrupted families, and so on [9–11]. In addition, HRQOL measures are also used to evaluate health service needs, and thereby influence public policy decisions such that early prevention and, if necessary, reconfiguration of health resources [12]. Therefore, the HRQOL in children and adolescents has attracted increased attention and has become an emerging area of research field [13].

Developing an appropriate instrument is vital for monitoring HRQOL. It is well known that the measurement of HRQOL in children and adolescents is different from that in adults. In addition to common dimensions such as physical and psychological well-being, it is important to realize the specific aspects of their families, friends, schools, and the communities. Furthermore, it is now widely endorsed that children and adolescents can answer HRQOL instruments reliably if their emotional development, cognitive capacity, and reading skills are taken into accounts [8]. Thus, self-report is an ideal way to obtain their subjective perspective. Finally, because HRQOL is extremely susceptible to cultural backgrounds, the instrument should be developed simultaneously among different cultures, or the existing accepted instrument should be translated, analyzed, and adapted into other languages considering the cultural diversities [14].

The KIDSCREEN instrument is a typical example of cross-cultural applications. The KIDSCREEN project was promoted by the European Commission and aimed to produce a generic instrument to assess HRQOL for children and adolescents aged between 8 and 18 years [15]. Compared with other instruments, KIDSCREEN has some unique advantages [10]: Firstly, it has excellent cross-cultural comparability because it was developed and tested simultaneously in several different European countries. Secondly, based on a broad perspective on the HRQOL of children and adolescents, it not only covers physical, psychological, and social aspects, but also creatively adds other unique new aspects, such as self-perception, autonomy, bullying, and financial resources. Thirdly, classical test theory (CTT) and item response theory (IRT) were combined in the process of development. Finally, there are three KIDSCREEN versions (KIDSCREEN-52, KIDSCREEN-27, and KIDSCREEN-10), and all have available corresponding rating as child/adolescent and parent/proxy rating.

Up to date, the KIDSCREEN has been successfully translated into more than 30 different languages in Europe, North and Latin America, Africa, Asia, and Oceania [16]. Besides, its good psychometric properties were also confirmed in a sample of Hongkong children and adolescents [17], and the

Cantonese Chinese version was formed. However, to our best knowledge, there is no Mandarin Chinese version of the KIDSCREEN. It is well known that Mandarin is the common language in Chinese mainland, and was adopted as the official language of the People's Republic of China (PRC) in 1955 [18]. Thus, its function is positioned as a standard language throughout China [19]. In contrast, Cantonese is spoken primarily in the southern provinces of Guangdong and Guangxi, and in the Hong Kong and Macao Special Administrative Regions of China [20]. Although Mandarin and Cantonese come from the same language group (Sino-Tibetan languages), they are quite different in terms of vocabulary, grammar structure, character pattern (simplified Chinese vs. traditional Chinese), and so on [18, 21, 22]. Hence, it is necessary to validate the Mandarin Chinese version of KIDSCREEN. The aim of this study was to evaluate the psychometric properties of the Mandarin Chinese version of KIDSCREEN-52.

Methods

Recruitment and data collection

The cross-sectional study was divided into two cohorts (baseline and test–retest survey) and was conducted in Weifang (a central section city in Shandong Province of China), carrying about 9.27 million inhabitants (9.42% of the total Shandong population) from October 2016 to November 2016. Weifang is divided into nine urban areas based on geographic location, three of which (Kuiwen, Weicheng, and High-tech Development) are located in the center. These three areas also represent different socioeconomic backgrounds, population size, and so forth, respectively. The basic school education in these three areas is coeducation and is divided into the following cycles: Preschool (3-year cycle, aged about 3–5); Primary school (6-year cycle, aged about 6–11); Middle school (3-year cycle, aged about 12–14); and High school (3-year cycle, aged about 15–18). In this study, considering the age range of the KIDSCREEN-52, only adolescents from the sixth grade in primary school to the second grade in high school (a total of six grades) were eligible to enter the research.

In the baseline survey, we conducted a multistage, stratified, and cluster sampling scheme that used the school-class as a sampling unit, stratified by the type of area, school and grade, and randomly selected at the school and class level. Firstly, according to the official records (Weifang Statistical Yearbook, 2016) and the website of the Weifang Education Bureau, we created a list of primary schools, middle schools, and high schools in each of three urban areas. Then an original sampling frame was formed. Secondly, we used various ways (such as official records, contact principals, and so on)

to screen all schools, and to identify all potential schools that fit for the study criteria (for example the number of students, school type, school size, and so forth). Thus, the resulting sampling was framed. Thirdly, we randomly selected schools from each urban area, and the number of primary, middle, and high schools was four, two, and one, respectively. Two nine-year coherent schools, which included both primary and middle within the same school, had been sampled within one area, so there were five schools in this area and seven in the rest of each two areas. Thus, a total of 19 schools were included in this study. In addition, schools that did not agree to participate were replaced by another randomly selected school from the same group. Finally, according to the research criteria (age, sex, etc.), we randomly selected classes from each grade, and all 4693 potential participants from 100 classes were included in the study.

The data were collected in the school environment by a pen and paper self-report questionnaire. Firstly, before the investigation of every class, sufficient questionnaires and a brief questionnaire guide were all put in a large sealed envelope by trained research assistants. Secondly, at the time of the adolescents' self-study in the afternoon, each class adviser took the sealed envelope to their own class and then opened it on the spot and distributed the questionnaires. Thirdly, the class advisers read the questionnaire guide aloud and provided the brief introduction and notes about the questionnaire. Fourthly, the questionnaire was completed under quiet classroom conditions within the given 35-min period. After doing so, they were told to continue with their school work to minimize any noise until all the other adolescents finished. In addition, they were allowed to ask research assistants for help if they did not understand any of the questions. Finally, research assistants collected all the questionnaires and put them back in the previous large envelope, re-sealed and took it away. The inclusion criteria for the adolescents were as follows: (1) age between 11 and 17 years; (2) able to read and complete the questionnaires independently; (3) consent to participate in the study; (4) at school on the day of the data collection. In the end, we screened 4385 eligible participants from 4693 potential participants, for a response rate of 93.4%.

After two weeks of the baseline survey, the test–retest samples, including 19 classes from 11 schools in three areas, were selected by the convenient and cluster method. Student identity numbers of potential respondents were used to allow the matching of test–retest questionnaires. Before the test–retest survey started, the class advisers of all selected classes were asked to provide details about potential participants of the last two weeks. These details include physical health, mental health, major life events, and so on, which might lead to significant changes in HRQOL. We did our best to ensure that the environment, time, and process of test–retest survey were highly similar to the baseline survey.

The inclusion criteria were as follows: (1) adolescents who involved in the baseline survey; (2) adolescents who completed the questionnaire in the baseline survey; (3) adolescents who agreed to participate in the test–retest survey; (4) adolescents who did not experience significant physical or mental changes within two weeks after the baseline survey; (5) adolescents who did not experience major life events within two weeks after the baseline survey. In the end, 841 qualified participants were selected from 867 potential participants and participated in the test–retest survey, for a response rate of 97.0%.

Ethical considerations and informed consent

This study was conducted after approval by the Ethics Committee of Weifang Medical University. Initially, the principals and school doctors from all participating schools were contacted for approval of participation and gave their permission. Next, each class adviser sent out and collected the parents' informed consent through the parent conference or the We-Chat group. Finally, the informed consent was also arranged on the first page of the questionnaire and in the teacher's guide manual. Participation was voluntary, and each adolescent decided whether to fill in the questionnaire after reading the informed consent. It is worth noting that, if adolescents want to quit after the study begins, they can raise their hands and stop answering without having to explain why.

Measures

KIDSCREEN-52 (Mandarin Chinese version)

The English original KIDSCREEN-52 is available in two versions as a self-report and parent/proxy report. In this study, only responses from adolescents on the KIDSCREEN-52 were used. The KIDSCREEN-52 self-report version, consisting 52 items, provides detailed information which are used to assess HRQOL across ten dimensions: physical well-being (5 items), psychological well-being (6 items), moods and emotions (7 items), self-perception (5 items), autonomy (5 items), parent relation and home life (6 items), financial resources (3 items), social support and peers (6 items), school environment (6 items), and social acceptance and bullying (3 items). The time frame refers to the previous week. Responses are categorized into on a five-point Likert scale that evaluates the frequency (1 = never, 2 = seldom, 3 = sometimes, 4 = often, 5 = always) or the intensity (1 = not at all, 2 = slightly, 3 = moderately, 4 = very, 5 = extremely) of certain behaviors, feelings and, attitudes. Negatively formulated items are reverted, and item scores for each respective dimension are summed up. The scores for each of the ten dimensions are transformed into T-values

with a mean of 50 and a standard deviation (SD) of 10, based on a representative sample of the European general children and adolescents [10]. Higher scores indicate better HRQOL.

Our research group obtained permission from the European KIDSCREEN Group to use the KIDSCREEN-52 self-report version. Based on the English original version of KIDSCREEN-52, a Mandarin Chinese translation was conducted. The translation process used the forward-backward-forward methodology [10, 23, 24]. Firstly, two translators, both native Mandarin Chinese speakers and fluent in English translated the KIDSCREEN-52 into Mandarin Chinese independently. Secondly, all the items of forward translations were then compared and assessed by a research group (including two forward translators and experts) in order to generate a single Mandarin Chinese reconciled version. Thirdly, two other professional bilingual translators back translated the Mandarin Chinese reconciled version into English separately. It was worth noting that the two backward translators were not familiar with the original questionnaire and did not participate in the forward translation process. Fourthly, all the items of backward translations were then compared and assessed by another research group (including two backward translators and experts), and a revised English backward translation version was produced. Lastly, the small differences between the original version and the backward translated version, which were of a culture and linguistic nature, were then discussed in the integrated research group (all translators and experts) to improve the quality of the Mandarin Chinese translation and to reach consensus. The above process was repeated several times to determine the semantic equivalence of all items. In order to ensure the accuracy and readability of all the items, we recruited a number of experts from relevant fields in China and conducted two rounds of expert consultation by email. Then, we revised the items according to the experts' comments. Finally, we got the pre-final Mandarin Chinese version of the KIDSCREEN-52. Furthermore, to ensure the items of the pre-final Mandarin Chinese version were age-appropriate, we conducted a pilot study on a small group of 48 boys and girls between 11 and 14 years old, using convenience sampling method. They were asked to fill out the questionnaire and give feedback on the readability and applicability of all items. Then the phrasing of some items was further minor adjusted, with no items substantially modified, added, or removed. The time required to fill in the questionnaire was also recorded by the researchers.

PedsQL™ 4.0 (Chinese version)

The Pediatric Quality of Life Inventory (PedsQL™ 4.0) is a generic instrument for assessing HRQOL in health as well as chronically ill children and adolescents between the ages of 8 and 18 years [25]. It consists of 23 items and the following

four dimensions: physical function (8 items), emotional function (5 items), social function (5 items), and school function (5 items). Responses are categorized into a five-point Likert scale that evaluates the frequency (0 = never; 1 = seldom; 2 = sometimes; 3 = often; 4 = always), with a one-month recall period. The dimension score was then converted into a 0–100-point scale, with 100 indicating the best HRQOL and 0 the worst [26]. Higher scores indicate better HRQOL. The Chinese version of the self-report PedsQL™ 4.0 has been validated in previous studies [27, 28]. In this study, it was used to evaluate the convergent and discriminant validity of the KIDSCREEN-52 and the Cronbach's alpha coefficient was satisfactory at 0.929.

FAS (Chinese version)

The socioeconomic status of the adolescent's families is assessed with the Family Affluence Scale (FAS), which includes questions on family car ownership, having own unshared room, the number of computers at home, and how many times the family spent on holidays in the past 12 months [29]. The FAS scores are collected in eight categories ranging from 0 to 7, which are recorded into three broad groups in the analysis (low 0–3, intermediate 4–5, and high 6–7) [30, 31]. The cross-cultural validity of the FAS has been shown in multinational surveys across over 30 countries [30] and the Chinese version has been validated in a previous study [31]. In this study, the Cronbach's alpha coefficient was moderate at 0.591. Higher family affluence was expected to be associated with higher scores in all KIDSCREEN-52 dimensions, but especially for the financial resources dimension.

SDQ (Chinese version)

The Strength and Difficulties Questionnaire (SDQ) is a brief screening questionnaire designed to evaluate mental health and behavioral difficulties of children and adolescents aged from 4 to 16 years [32]. It exists in three informant versions (parent report, teacher report, and self-report), with each version consists of 25 items measuring five dimensions of emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior. Each dimension contains five items. Responses are categorized into a three-point scale (0 = not true, 1 = somewhat true, and 2 = certainly true), with a six-month recall period. A small number of items are reversed scored. A total difficulties score is obtained by summing up the scores for all except the Prosocial Behavior dimension. The possible range of scores for each of the dimensions is 0–10, and for the total difficulty score is 0–40. Thus, the total difficulty score can be regarded as normal (0–14), borderline (15–17), and abnormal (18–40) by cutoff values derived from the Shanghai

norm data [33]. Higher scores indicate more problems. Considerable research has been conducted in numerous Western community contexts to establish the psychometric properties of the SDQ [34–36] and the Chinese version has also demonstrated robust reliability and validity not only in Chinese mainland [33], but also in Hongkong [37] and Taiwan [38]. In this study, the self-report version of SDQ was adopted, and the Cronbach's alpha coefficient was satisfactory at 0.871. Higher difficulties scores were hypothesized to be associated with lower scores in all KIDSCREEN-52 dimensions, but especially in the mood and emotions and psychological well-being dimensions.

Socio-demographic characteristics

A socio-demographic information sheet covered basic participant information, including age, gender, ethnicity, siblings, parental education level, and socioeconomic status.

Statistical analysis

Feasibility

The feasibility of the KIDSCREEN-52 was determined by the response rate and the percentage of the missing values in each dimension.

Item and dimension properties

As the measures of item properties, corrected item-domain correlations coefficients and "Cronbach's alpha if item deleted" were calculated [39, 40]. The floor and ceiling effects of all dimensions were also calculated. If more than 15% of respondents achieved the lowest or highest possible scores on a given dimension, they were defined as being present [41]. In addition, following the KIDSCREEN Group Europe's scoring algorithm [10], Rasch scores were computed for each dimension, weighted in accordance with the European norm population, and reported as T-values, with a mean of 50 and a standard deviation of 10.

Reliability

Reliability was determined by internal consistency using Cronbach's alpha coefficient and test–retest reliability using intraclass correlation coefficients (ICCs). Alpha coefficients > 0.70 indicate acceptable reliability, whereas that > 0.90 is recommended for better [42]. The ICC values for the levels of agreement consider < 0.20 as poor agreement, $0.21–0.40$ as fair, $0.41–0.60$ as moderate, $0.61–0.80$ as good, and > 0.81 as very good [43].

Construct validity

The construct validity of the KIDSCREEN-52 was tested using confirmatory factor analysis (CFA) with maximum likelihood estimation (MLE). The aim of CFA was to test the hypothesis that there existed a relationship between the observed variables (items) and their underlying latent constructs (dimensions). The overall model fit was assessed using the Chi-square test and ratio of Chi-square to degrees of freedom (χ^2/df). Reference values indicate a good model with $\chi^2/df < 2$ and acceptable with < 3 [44]. However, as this test is highly sensitive to sample size, we used alternative fit indices having the following cutoff values suggesting acceptable fit. Comparative fit index (CFI), normed fit index (NFI), Tucker-Lewis index (TLI), and adjusted goodness-of-fit index (AGFI) were reported with values of > 0.90 [45] or > 0.95 [46], which indicate acceptable or good model fit, respectively. The standardized root mean residual (SRMR) values < 0.05 suggesting good model fit [44]. In addition, the root mean squared error of approximation (RMSEA) was reported as an absolute fit index, with values < 0.08 or 0.05 considered acceptable or good fit, respectively [46].

Measurement invariance

The multigroup confirmatory factor analysis was performed to test the measurement invariance across gender and age groups in the whole sample. For testing loading invariance, $\Delta CFI \geq -0.010$, supplemented by $\Delta RMSEA \geq -0.015$ or $\Delta SRMR \geq -0.030$, would indicate non-invariance [47]. In this study, the measurement invariance of the KIDSCREEN-52 was tested for configural equivalence, and results were used as the baseline model for subsequent analyses. After the metric invariance was obtained, scalar invariance and strict invariance were tested.

Convergent and discriminant validity

The convergent and discriminant validity were further assessed by determining the degree of correlation between two instruments assessing similar or less comparable dimensions of HRQOL. In this study, convergence indicates the two dimensions (such as KIDSCREEN-52's physical well-being and PedsQL™ 4.0's physical function) believed to reflect the same underlying concept highly correlated each other, whereas discriminant validity indicates low or moderate correlations between dimensions (e.g., KIDSCREEN-52's physical well-being and PedsQL™ 4.0's school function) that are believed to assess different characteristics. Pearson correlation coefficient (r) was used to measure the strengths of the association, and $r < 0.1$ was considered trivial, $0.10–0.30$ as low, $0.31–0.50$ as moderate, and ≥ 0.5 as high [41].

Known-group validity

The known-group validity of the KIDSCREEN-52 was examined by comparing the results between groups which were a priori expected to show differences in HRQOL. We conducted a series of one-way analyses of variance (ANOVA) to test the significant differences between socio-economic status (low/medium/high) and mental health status (abnormal/borderline/normal) in relation to all dimensions. In addition, two kinds of effect sizes (η_p^2 and Cohen's d) were calculated. Partial eta squared (η_p^2), which was considered to be suitable for ANOVA, was able to indicate the strength of the effect of group differences in all dimensions. However, in the original study of the KIDSCREEN, Cohen's d was used as a measure of effect size, by dividing the difference between the groups (e.g., low FAS vs. high FAS) on each of the dimension [10]. Thus, in this study, Cohen's d was also calculated to compare with the original research. Following conventions outlined [48], we interpreted effect size magnitudes of η_p^2 (0.01 as small, 0.06 as moderate, and 0.14 as large) and Cohen's d (0.20 as small, 0.50 as moderate, and 0.80 as large), respectively.

Missing values

Prior to running our analyses, we made decisions about missing values. The number of items with missing values was low. All 52 items of the KIDSCREEN were completed in 3900 participants (88.94%), without any missing values. As a structural equation model (SEM) is unsuitable for handling datasets that do contain missing values, we used 3900 complete data sets for confirmatory factor analysis. In addition, missing values in other calculations dealt with by the pairwise exclusion option.

Epiadta3.1 (EpiData Association, Odense, Denmark) was used to establish the database. Cohen's d was calculated by the G*Power 3.1 computer program [49]. In addition, confirmatory factor analysis was performed by AMOS 24.0, and all remaining data analyses with SPSS 24.0 for Windows (SPSS, Inc., Chicago, IL). A p value of less than 0.05 was accepted as statistically significant.

Results

Socio-demographic characteristics

A sample of 4385 adolescents participated at baseline and 841 at test–retest surveys. In both surveys, the average age of adolescents was 13.71 and 13.81 years old, respectively, with a range from 11 to 17 years. The sex ratio of male and female was roughly the same (46.9%, 53.1% vs. 48.5%, 51.5%). The overwhelming majority of adolescents (99.1%, 99.5%) were the Han ethnicity and slightly less than half (48.7%, 47.8%) came from the one-child family. The majority of parental education level was secondary school, reaching (65.3%, 68.4%) and (68.2%, 69.8%), respectively. In addition, 41.5% and 34.8% of adolescents were from families with high socioeconomic status (Table 1).

Feasibility

The response rates of baseline and test–retest survey were 93.4% and 97.0%, respectively. In addition, the percentage of missing values were very low, ranging from 0.02% for social acceptance and bullying to 1.92% for physical well-being of all dimensions (Table 2).

Table 1 Socio-demographic characteristics of the baseline and test–retest sample

Characteristics	Baseline sample ($n=4385$)	Test–retest sample ($n=841$)
Age [mean (SD)]	13.71 (1.864)	13.81 (1.877)
Gender [n (%)]		
Male	2056 (46.9)	408 (48.5)
Female	2329 (53.1)	433 (51.5)
Ethnicity [n (%)]		
Han ethnicity	4345 (99.1)	837 (99.5)
Ethnic minority	40 (0.9)	4 (0.5)
Siblings [n (%)]		
None	2135 (48.7)	402 (47.8)
One	2043 (46.6)	412 (49.0)
Two or more	207 (4.7)	27 (3.2)
Father's education level [n (%)]		
Primary school or less	167 (3.8)	41 (4.9)
Middle school	1522 (34.7)	323 (38.4)
High school	1344 (30.6)	251 (29.8)
Junior college	781 (17.8)	123 (14.9)
College/university or higher	571 (13.0)	101 (12.0)
Mother's education level [n (%)]		
Primary school or less	352 (8.0)	80 (9.5)
Middle school	1722 (39.3)	360 (42.8)
High school	1278 (29.1)	227 (27.0)
Junior college	665 (15.2)	101 (12.0)
College/university or higher	368 (8.4)	73 (8.7)
Socioeconomic status [n (%)][a]		
Low FAS	791 (18.0)	178 (21.2)
Medium FAS	1775 (40.5)	370 (44.0)
High FAS	1819 (41.5)	293 (34.8)

[a] *FAS* Family Affluence Scale (0–3 = low; 4–5 = medium; 6–7 = high)

Table 2 Description, item and dimension properties, internal consistency, and test–retest reliability of the KIDSCREEN-52

	Items	n^a	Missing (%)	Floor (%)	Ceiling (%)	Mean T-value (SD)	α	α if item deleted	Corrected item-domain correlations	n^b	ICC [95% CI]
KIDSCREEN-52 dimensions											
Physical well-being	5	4301	1.92	0.02	7.42	51.41 (8.02)	0.854	0.819–0.830	0.645–0.690	812	0.800 [0.763, 0.828]
Psychological well-being	6	4339	1.05	0.02	12.88	47.94 (9.77)	0.931	0.913–0.923	0.761–0.841	828	0.836 [0.814, 0.856]
Moods and emotions	7	4331	1.23	0.02	12.63	47.57 (12.12)	0.958	0.948–0.954	0.819–0.889	825	0.845 [0.826, 0.863]
Self-perception	5	4354	0.71	0.14	10.45	48.39 (8.50)	0.877	0.844–0.863	0.656–0.733	823	0.760 [0.719, 0.791]
Autonomy	5	4342	0.98	0.12	15.50	48.43 (11.71)	0.959	0.947–0.954	0.857–0.904	816	0.832 [0.812, 0.850]
Parent relation and home life	6	4345	0.91	0.02	14.87	47.19 (9.03)	0.906	0.879–0.905	0.622–0.807	816	0.819 [0.792, 0.842]
Financial resources	3	4320	1.48	0.37	17.27	45.12 (9.24)	0.919	0.866–0.909	0.806–0.856	822	0.809 [0.791, 0.826]
Social support and peers	6	4337	1.09	0.05	8.55	48.55 (8.62)	0.908	0.886–0.897	0.714–0.796	816	0.832 [0.807, 0.856]
School environment	6	4342	0.05	0.98	6.75	49.46 (8.44)	0.917	0.898–0.907	0.729–0.794	815	0.849 [0.829, 0.866]
Social acceptance and bullying	3	4325	0.02	1.37	31.84	46.64 (9.65)	0.819	0.724–0.777	0.649–0.705	821	0.724 [0.681, 0.766]

All correlations are significant at the 0.01 level (two-tailed)

α, Cronbach's alpha

[a]Number of responders of the baseline survey

[b]Number of responders of the test–retest survey

Item and dimension analysis

The corrected item-domain correlations coefficients for all the 52 items of KIDSCREEN-52 were 0.645–0.690, 0.761–0.841, 0.819–0.889, 0.656–0.733, 0.857–0.904, 0.622–0.807, 0.806–0.856, 0.714–0.796, 0.729–0.794, and 0.649–0.705, respectively ($p_s < 0.01$). In addition, dropping either item in the belonged dimension had a small negative effect on respective Cronbach's alpha coefficients. Furthermore, deleting either item had a smaller effect on the total Cronbach's alpha coefficients, with alpha of the remaining 51 items remaining at 0.963 or 0.964. The mean scores varied from 45.12 (SD = 9.24) for financial resources to 51.41 (SD = 8.02) for physical well-being due to *T*-value standardization. The floor effects ranged from 0.02% for (physical well-being, psychological well-being, moods and emotions, and parent relation and home life) to 1.37% for social acceptance and bullying, while ceiling effects ranged from 6.75% for school environment to 31.84% for social acceptance and bullying (Table 2).

Internal consistency and test–retest reliability

The Cronbach's alpha coefficient for the total score was 0.964, and the alpha coefficient values for all dimensions ranged from 0.819 for social acceptance and bullying to 0.959 for autonomy. The test–retest reliability was assessed in a sub-sample of 841 adolescents, and the interval of test–retest measurement was approximately two weeks. The ICCs between scale scores for the two assessments ranged from 0.724, 95% CI [0.681, 0.766] for social acceptance and bullying to 0.849, 95% CI [0.829, 0.866] for school environment ($p_s < 0.01$). Details are shown in Table 2.

Construct validity

We further evaluated the ten-factor model by using the confirmatory factor analysis. The goodness-of-fit statistics of the ten-dimensional model based on the original KIDSCREEN-52 structure were also calculated. The $\chi^2_{(df=1229)} = 8500.4$ ($p < 0.001$) and $\chi^2/df = 6.917$, which falls outside the acceptable range. However, other fit indices supported the ten-factor structure: CFI = 0.955, NFI = 0.947, TLI = 0.951, AGFI = 0.901, SRMR = 0.033 and RMSEA = 0.039, 90% CI [0.038, 0.040].

Measurement invariance

Results regarding the measurement invariance of the ten dimensions of KIDSCREEN-52 across gender and age groups are presented in Table 3. The results of four steps from loose to strict suggested its invariance across gender: ΔCFI = 0.000, − 0.001, and − 0.001 < − 0.01; ΔRMSEA = 0.000, 0.000, and 0.000 < 0.015. However, in consideration of age groups, the KIDSCREEN-52 cannot be considered fully invariant. The results of ΔRMSEA (0.001, 0.002, and 0.004 < 0.015) suggested its invariance, while ΔCFI of scalar and strict invariance was − 0.012 and − 0.022 > − 0.01.

Convergent and discriminant validity

The correlations between KIDSCREEN-52's dimensions and PedsQL™ 4.0's dimensions are presented in Table 4. The KIDSCREEN-52's dimension scores were positively and significantly correlated with the dimensions of PedsQL™ 4.0. As expected, the correlations of the two dimensions dealing with similar structures displayed a high level of correlation, suggesting the convergent validity: the

Table 3 Measurement invariance of the KIDSCREEN-52 across gender and age groups

Models	Model fit indices								
	χ^2	*df*	χ^2/df	CFI	RMSEA [90% CI]	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA
Gender groups									
Model 1: Configural invariance	9962.4***	2458	4.053	0.953	0.028 [0.027, 0.029]				
Model 2: Metric invariance	10033.8***	2500	4.014	0.953	0.028 [0.027, 0.028]	71.4	42	0.000	0.000
Model 3: Scalar invariance	10191.2***	2555	3.989	0.952	0.028 [0.027, 0.028]	228.8***	97	− 0.001	0.000
Model 4: Strict invariance	10327.9***	2607	3.962	0.952	0.028 [0.027, 0.028]	365.5***	149	− 0.001	0.000
Age groups									
Model 1: Configural invariance	11571.7***	3687	3.139	0.949	0.023 [0.023, 0.024]				
Model 2: Metric invariance	12406.0***	3771	3.290	0.944	0.024 [0.024, 0.025]	834.3***	84	− 0.005	0.001
Model 3: Scalar invariance	13683.4***	3881	3.526	0.937	0.025 [0.025, 0.026]	2111.7***	194	− 0.012	0.002
Model 4: Strict invariance	15397.6***	3985	3.864	0.927	0.027 [0.027, 0.028]	3825.9***	298	− 0.022	0.004

df, degree of freedom; RMSEA, root mean square error of approximation; CFI, comparative fit index

***Significant at $p < 0.001$

Table 4 Pearson correlation coefficients between the dimensions of KIDSCREEN-52 and PedsQL™ 4.0

	PedsQL™ 4.0 dimensions			
	Physical functioning	Emotional functioning	Social functioning	School functioning
KIDSCREEN-52 dimensions				
Physical well-being	**0.705****	0.395**	0.365**	0.456**
Psychological well-being	0.362**	0.587**	0.497**	0.491**
Mood and emotions	0.389**	**0.894****	0.556**	0.598**
Self-perception	0.350**	0.579**	0.468**	0.502**
Autonomy	0.232**	0.449**	0.347**	0.366**
Parent relation and home life	0.289**	0.483**	0.441**	0.437**
Financial resources	0.230**	0.331**	0.311**	0.376**
Social support and peers	0.400**	0.509**	**0.725****	0.556**
School environment	0.347**	0.594**	0.553**	**0.764****
Social acceptance and bullying	0.376**	0.501**	0.703**	0.546**

r, Pearson correlation coefficient; PedsQL™ 4.0, Pediatric Quality of Life Inventory

Values in bold showed the highest correlations of each column

**Significant at $p < 0.01$

PedsQL™ 4.0's emotional functioning showed the highest correlation with the KIDSCREEN-52's mood and emotions ($r = 0.894$), followed by that between school functioning and school environment ($r = 0.764$), social functioning and social support and peers ($r = 0.725$), physical functioning and physical well-being ($r = 0.705$). Conversely, the correlations between less comparable dimensions were low (e.g., 0.232, 0.289, 0.230 between KIDSCREEN-52's autonomy, parent relation and home life, and financial resources and PedsQL™ 4.0's physical functioning), which reveals the discriminant validity.

Known-group validity

Table 5 presents a gradient for all the KIDSCREEN-52's dimensions when the FAS is used to assess the socioeconomic status category of the family. A series of one-way ANOVA showed that KIDSCREEN-52's dimension scores on FAS were all significant ($F = 40.799$–1275.525, $p_s < 0.001$). In addition, two standardized effect sizes (η_p^2 and Cohen's d) were shown in the last two columns. A large $\eta_p^2 = 0.371$ was found in financial resources between these three

Table 5 Differences in KIDSCREEN-52's dimension scores by socioeconomic categories

	FAS				
	Low	Medium	High	Effect size	
	Mean T-value (SD)	Mean T-value (SD)	Mean T-value (SD)	η_p^2	Cohen's d (low vs. high)
KIDSCREEN-52 dimensions					
Physical well-being	48.96 (7.19)	50.64 (7.86)	53.21 (8.12)	0.042	**0.554**
Psychological well-being	45.10 (9.04)	46.91 (9.42)	50.16 (9.93)	0.041	**0.533**
Mood and emotions	43.18 (10.65)	45.94 (11.69)	51.08 (12.19)	**0.066**	**0.690**
Self-perception	46.23 (7.32)	47.33 (7.96)	50.36 (9.07)	0.040	**0.501**
Autonomy	45.40 (10.38)	47.34 (11.29)	50.81 (12.20)	0.033	0.478
Parent relation and home life	45.03 (8.43)	46.90 (9.32)	48.42 (8.80)	0.018	0.393
Financial resources	38.18 (5.01)	41.46 (5.90)	51.62 (9.17)	**0.371**	**1.819**
Social support and peers	45.45 (6.91)	47.42 (7.92)	50.99 (9.27)	**0.063**	**0.678**
School environment	46.35 (6.29)	48.21 (7.78)	52.03 (9.11)	**0.072**	**0.726**
Social acceptance and bullying	43.10 (8.61)	45.33 (9.40)	49.46 (9.55)	**0.067**	**0.700**

FAS Family Affluence Scale

η_p^2: 0.01 = small, 0.06 = medium, 0.14 = large; Cohen's d: 0.20 = small, 0.50 = medium, 0.80 = large

The effect size of "medium" or greater is indicated in bold

groups (low/medium/high) and η_p^2 in the mood and emotions, social support and peers, school environment, and social acceptance and bullying were medium (0.063–0.072). η_p^2 of remaining dimensions were all small (0.018–0.042). Regarding the Cohen's d between two FAS groups (low vs. high), parent relation and home life and autonomy were small (0.393, 0.478). In addition, large Cohen's $d = 1.819$ was in financial resources, while medium Cohen's d were in the other dimensions (0.501–0.726).

Furthermore, SDQ problem scores were used to differentiate the mental health status between abnormal, borderline, and normal. Significant differences exist in all dimensions of KIDSCREEN-52 ($F = 193.589$–1277.164, $p_\text{s} < 0.001$). The η_p^2 in the physical well-being, autonomy, parent relation and home life, and financial resources were medium (0.087–0.136), while remaining dimensions were all large (0.162–0.385). Regarding the Cohen's d between two SDQ groups (abnormal vs. normal), Cohen's d were large (0.941–2.085) in all dimensions except that physical well-being and financial resources were medium (0.732, 0.738). Details are shown in Table 6.

Discussion

As an instrument for measuring HRQOL in children and adolescents, KIDSCREEN-52 can be widely used in epidemiologic studies, clinical intervention studies, and research projects [10]. To our best knowledge, this is the first time that the psychometric properties of the KIDSCREEN-52 are being evaluated on Chinese mainland adolescents. Overall, this study developed the Mandarin Chinese version of KID-SCREEN-52 and provided evidence that it has sufficient psychometric properties to assess HRQOL in adolescents. These findings will contribute to the global use of this instrument, especially for both country and culture comparisons.

This instrument had strong feasibility and acceptability, with convenient filling out, high response rates, and few missing values. The high response rates and few missing values were also appeared in previous international studies [50–53] on school sampling and administration. All these may be due to the following facts: Before the study, we obtained the informed consent from adolescents and their parents and gained a greater degree of trust among parents [50]. Then, in order to minimize disturbance to the survey, the scope of class advisers was limited to outside the classroom, after completing a series of tasks such as leading research assistants into their class, briefly introducing the research content and helping to distribute the questionnaire. In addition, adolescents were also informed that only the researcher will access their questionnaire and read their answers. And, research assistants were on the spot to help if necessary [54]. In a word, the instrument might be viable in a cross-cultural setting.

Excellent psychometric properties were demonstrated in all items and dimensions of this instrument. The corrected item-domain correlations coefficients of all items were all greater than 0.60. In addition, the negative effects of dropping either item on the Cronbach's alpha coefficient were negligible, regardless of belonged dimension or total score. We found almost negligible floor effects, but a slight ceiling

Table 6 Differences in KIDSCREEN-52's dimension scores by mental health status

	SDQ				
	Abnormal	Borderline	Normal	Effect size	
	Mean T-value (SD)	Mean T-value (SD)	Mean T-value (SD)	η_p^2	Cohen's d (abnormal vs. normal)
KIDSCREEN-52 dimensions					
Physical well-being	47.40 (7.50)	48.66 (6.30)	53.03 (7.87)	**0.093**	**0.732**
Psychological well-being	40.80 (7.59)	43.73 (7.13)	50.77 (9.32)	**0.191**	**1.173**
Mood and emotions	35.25 (4.21)	38.63 (6.20)	52.57 (10.97)	**0.385**	**2.085**
Self-perception	42.70 (5.49)	44.90 (6.29)	50.63 (8.52)	**0.162**	**1.106**
Autonomy	41.26 (9.66)	43.80 (9.23)	51.21 (11.42)	**0.135**	**0.941**
Parent relation and home life	41.49 (7.53)	44.58 (8.00)	49.39 (8.72)	**0.136**	**0.970**
Financial resources	40.72 (7.16)	41.87 (7.24)	46.96 (9.57)	**0.087**	**0.738**
Social support and peers	41.92 (6.55)	47.30 (6.90)	50.84 (8.23)	**0.183**	**1.199**
School environment	42.35 (5.26)	45.68 (4.98)	52.20 (8.11)	**0.246**	**1.441**
Social acceptance and bullying	38.44 (7.77)	43.86 (7.50)	49.59 (8.80)	**0.230**	**1.343**

SDQ Strength and Difficulties Questionnaire

η_p^2: 0.01 = small, 0.06 = medium, 0.14 = large; Cohen's d: 0.20 = small, 0.50 = medium, 0.80 = large

The effect size of "medium" or greater is indicated in bold

effect in the autonomy and financial resources dimensions, and a moderate in the social acceptance and bullying dimension. Similar findings have been described in the original study [10] and further, the Mandarin Chinese version actually reduces the ceiling effect to some extent compared to the original study. Studies have shown that the ceiling effect may be expected in general HRQOL instruments because they are designed to be suitable for a wide range of populations [55]. Therefore, this study is aimed at general adolescent samples, where higher ceiling effects would be expected. Further testing of this version is required in clinical samples, where the ceiling effect would likely be significantly reduced.

The internal consistency of the Mandarin Chinese version was appropriate for all of ten dimensions, which similar to the original result of the KIDSCREEN project [16]. In other recent studies, excellent internal consistency of KIDSCREEN-52 in adolescents was also demonstrated [14, 17, 51, 56]. The test–retest ICCs were slightly higher than the a priori defined threshold of 0.70 [41] and the range of 0.560–0.770 reported previously [10]. This might be attributable to the following reasons: The maximum degree of informed consent and cooperation were obtained through convenience sampling in test–retest survey. Strict sample matching was performed to eliminate confounding factors that might influence the dramatic changes in HRQOL of adolescents. We conducted both surveys of school-based sampling and management, which maximized the similarity of environment, time, and process. Therefore, internal consistency and test–retest reliability were adequate for the Mandarin Chinese version.

The confirmatory factor analysis yielded support for the ten-dimensional factor model with results that are consistent with other recent European, Asian, and African studies [51, 56–59]. Our results are similar to the previous studies on adolescents with the CFI, NFI, TLI, and AGFI values greater than 0.90 [51, 56, 58–60] and RMSEA value less than 0.05 [57, 61]. The significant Chi square was observed, which might indicate a misfit of the ten-dimensional structure. However, this is more likely to be due to the fact that the Chi-square statistic alone is not appropriate in large data sets [62] and this situation has also appeared in previous studies [10, 56–59, 61]. Additionally, our findings supported the full configure invariance and metric invariance of the ten-dimensional factor model of gender and age groups. However, at a more stringent level (scalar and strict invariance), the age group cannot be considered fully invariant. To the best of our knowledge, this study is the first to test the measurement invariance at a more stringent level in the Mandarin Chinese version of KIDSCREEN-52 across gender and age groups, and the cause of this result is not yet clear. Future use of this instrument with adolescents should test for the measurement invariance [63] and, if necessary, identify non-invariant items by constraining the intercepts

of each item individually [64] and removed or unconstrained these items [65].

The convergent and discriminant validity analysis indicated that the Mandarin Chinese version showed a reasonable pattern of associations. We analyzed the validity by correlating the KIDSCREEN-52 with a similar instrument (PedsQL™ 4.0) to assess HRQOL. Correlations were generally the highest for those pairs of dimensions in which higher correlations were a priori expected. For example, the highest correlation was found between moods and emotions of KIDSCREEN-52 with emotional functioning of PedsQL™ 4.0. This was considered to reflect a reasonable convergence, while low to moderate correlations between theoretical different dimensions support the construct validity in the form of discriminant validity (that is, the lowest correlation between the KIDSCREEN-52's financial resources with PedsQL™ 4.0's physical functioning was found). Previous studies have also shown support for the validity of the KIDSCREEN-52 by correlations between KIDSCREEN-52 and PedsQL™ 4.0 [52, 66].

All KIDSCREEN-52 dimensions were significantly differentiated according to the different socioeconomic status, especially in high vs. low. Although other previous studies have generally shown that the KIDSCREEN are capable of discriminating adolescents with different socioeconomic status [10], this study shows a more magnitude of the effect sizes than the original research. This is because sample selection of this study was only for adolescents (lack of child samples), as a previous study has shown that socioeconomic status might be more important for HRQOL in adolescents than children [67]. Actually, one would expect a comparison of KIDSCREEN-52 between children and adolescents, but unfortunately it was beyond the scope of this study.

Furthermore, we found that the Mandarin Chinese version of KIDSCREEN-52 was able to discriminate between mental health status of adolescents, which was consistent with previous researches done on European samples [10]. However, our study found that the effect sizes in many dimensions were higher, especially between adolescents with poor vs. good mental health. The amazing variation in the magnitude of effect sizes has been reported in previous studies [67] and might be because many of the KIDSCREEN dimensions focus more on mental and social heath. In addition, we only surveyed 11- to 17-year-old adolescents lacking children groups, which may exacerbate the effect sizes, for the situation of physically and mentally in adolescents was more changeable and unequal [68]. These results should be confirmed in further research by using this instrument in clinical settings.

Some limitations of this study should be addressed. The first limitation is that this study limited the age of adolescents between 11 and 17 years (from the sixth grade in primary school to the second grade in high school). It

is more usual to assess a wider age group between 8 and 18 years. The absence of a sample of adolescents in the 18-year-old age group can be explained. The vast majority of them are in the last year of high school, facing an important turning point in life–Chinese College Entrance Examination. Thus, learning has become a top priority, and schools have policies that discourage these adolescents from participating in social activities unrelated to learning. Therefore, based on the above-mentioned facts, this study did not include adolescents of this age group. Furthermore, we believe that if good psychometric properties of an instrument are demonstrated in a large sample of younger adolescents, it is possible that this will also be the case in older adolescents as they typically have better cognitive skills [69]. However, the lack of the children sample aged from 8 to 10 is indeed a significant limitation, which makes it impossible to know the measurement performance of the instrument at this age and thus limiting its use in turn. Thus, given the importance of this age group in the development and application of this instrument, we encourage future researchers to consider this in their future study. The second limitation is that data were only sampled from Weifang of China, due to the aims of the survey and the limited resources. Given the diversities across ethnic groups, geographical regions, socioeconomic levels, and other factors in Chinese mainland, we must pay attention to the interpretation of the study's findings, as these may not be generalizable to the Chinese populations. Therefore, future studies should use a larger and representative sample of children and adolescents in Chinese mainland to further examine the psychometric properties of this instrument. The third limitation is that only two variables (socioeconomic and mental health status) were included in the known-group validity analysis, which ignored the sensitivity of the instrument to other social determinants, such as social support, family status, school environment, and health care utilization. Future studies would benefit from the use of multiple indicators of variables in order to allow for the evaluation of the measurement and structural models. The fourth limitation of the study is its reliance on self-report data, which are often biased by social desirability. Future research will need to establish that adolescent responses are related to other sources of information such as teacher and parents reports. The fifth limitation is that the instrument was not tested in a clinical setting; therefore, it needs to be tested in a clinical setting where clinical diagnoses and information about the severity of conditions are available. Finally, because of the cross-sectional design of the study, it was impossible to test the sensitivity of the instrument to change. This should be tested in future studies which might focus on testing the sensitivity of the instrument to change within a randomized longitudinal intervention study with a control-group.

Conclusion

In conclusion, evidence from this study supported that the Mandarin Chinese version of KIDSCREEN-52 is a feasible, reliable, and valid instrument for measuring HRQOL of adolescents in Weifang, China. It is worth noting that this is not the first study that validates the instrument in a Chinese context since the Cantonese version already exists. However, this is the first study to validate the instrument in a Chinese mainland adolescent population. Despite the important limitations mentioned above, we believe that this study provides a starting point for using KIDSCREEN-52 in the HRQOL measurements among adolescents in Chinese mainland. Therefore, although having the positive initial outcomes, the evaluation of KIDSCREEN-52 should be an on-going process, by extending psychometric testing to characteristics not only assessed in this study, but also assessing its applicability and performance in other populations and regions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Ethical approval was granted by the Ethics Committee of Weifang Medical University in April 2016.

Informed consent Informed consent was provided by all participants including in the study. The collection and transfer of data was carried out according to strict security and data encryption.

References

1. The WHOQOL Group. (1995). The World Health Organization Quality of Life assessment (WHOQL): Position paper from the World Health Organization. *Social Science & Medicine, 41*(10), 1403–1409. https://doi.org/10.1016/0277-9536(95)00112-K.
2. Taylor, R. M., Gibson, F., & Franck, L. S. (2010). A concept analysis of health-related quality of life in young people with chronic illness. *Journal of Clinical Nursing*. 17(14), 1823–1833. https://doi.org/10.1111/j.1365-2702.2008.02379.x.
3. Cieza, A., & Stucki, G. (2005). Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF). *Quality of Life Research, 14*(5), 1225–1237. https://doi.org/10.1007/s11136-004-4773-0.
4. Eiser, C., & Morse, R. (2001). A review of measures of quality of life for children with chronic illness. *Archives of Disease in Childhood, 84*(3), 205–211. https://doi.org/10.1136/adc.84.3.205.
5. Gandek, B., & Jr, W. J. (1998). Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. International Quality of Life Assessment. *Journal of Clinical Epidemiology, 51*(11), 953–959. https://doi.org/10.1016/S0895-4356(98)00086-9.
6. Higginson, I. J., & Carr, A. J. (2001). Measuring quality of life: Using quality of life measures in the clinical setting. *BMJ, 322*(7297), 1297–1300. https://doi.org/10.1136/bmj.322.7297.1297.
7. Zumbo, B. D. (2003). *Healthy Days, health satisfaction and satisfaction with the overall quality of life. Essays on the quality of life*. Netherlands: Springer. https://doi.org/10.1007/978-94-017-0389-5_11.
8. Matza, L. S., Swensen, A. R., Flood, E. M., Secnik, K., & Leidy, N. K. (2004). Assessment of health-related quality of life in children: a review of conceptual, methodological, and regulatory issues. *Value in Health, 7*(1), 79–92. https://doi.org/10.1111/j.1524-4733.2004.71273.x.
9. Furusho, J., Kubagawa, T., Satoh, H., Shibata, R., Nemoto, Y., Matsuzaki, K., et al. (2006). Study of the kid-KINDL questionnaire scores for children with developmental disorders in normal classes and their parents. *No to Hattatsu, 38*(3), 183–186.
10. The KIDSCREEN Group Europe. (2006). *The KIDSCREEN questionnaires: Quality of life questionnaires for children and adolescents. Handbook*. Lengerich: Pabst Science Publishers.
11. Nakazawa, J. (2006). Children's Peer Relations and Social Behavior: A Review of Recent Research in Japan. *Annual Report of Educational Psychology in Japan 45*, 185–196. https://doi.org/10.5926/arepj1962.45.0_185.
12. Centers for Disease Control and Prevention. (2000). *Measuring Healthy Days: Population Assessment of Health-Related Quality of Life*. https://www.cdc.gov/hrqol/pdfs/mhd.pdf.Accessed April 13, 2018.
13. Solans, M., Pane, S., Estrada, M. D., Serra-Sutton, V., Berra, S., & Herdman, M., et al. (2008). Health-related quality of life measurement in children and adolescents: A systematic review of generic and disease-specific instruments. Value in Health. 11(4), 742–764. https://doi.org/10.1111/j.1524-4733.2007.00293.x.
14. Stevanovic, D., Tadic, I., Novakovic, T., Kisic-Tepavcevic, D., & Ravens-Sieberer, U. (2012). Evaluating the Serbian version of the KIDSCREEN quality-of-life questionnaires: Reliability,

15. validity, and agreement between children's and parents' ratings. *Quality of Life Research, 22*(7), 1729–1737. https://doi.org/10.1007/s11136-012-0286-4.
16. Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Duer, W., et al. (2005). KIDSCREEN-52 quality-of-life measure for children and adolescents. *Expert Review of Pharmaco-economics & Outcomes Research, 5*(3), 353–364. https://doi.org/10.1586/14737167.5.3.35.
16. Ravens-Sieberer, U., Herdman, M., Devine, J., Otto, C., Bullinger, M., Rose, M., et al. (2014). The European KIDSCREEN approach to measure quality of life and well-being in children: development, current application, and future advances. Quality of Life Research. 23(3), 791–803. https://doi.org/10.1007/s11136-013-0428-3.
17. Ng, J. Y., Burnett, A., Ha, A. S., & Sum, K. W. (2015). Psychometric properties of the Chinese (Cantonese) versions of the KIDSCREEN health-related quality of life questionnaire. *Quality of Life Research, 24*(10), 2415–2421. https://doi.org/10.1007/s11136-015-0973-z.
18. Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary development in Mandarin (Putonghua) and Cantonese. *Journal of Child Language*. 36(5), 1115–1144. https://doi.org/10.1017/S0305000908009185.
19. Wang, L., & Ladegaard, H. J. (2008). Language attitudes and gender in China: Perceptions and reported use of Putonghua and Cantonese in the Southern Province of Guangdong. *Language Awareness, 17*(1), 57–77. https://doi.org/10.2167/la425.0.
20. Matthews, S., & Yip, V. (2011). *Cantonese: A Comprehensive Grammar* (2nd edn.). London: Routledge.
21. Snow, D. (2004). *Cantonese as written language: The growth of a written Chinese vernacular*. Hong Kong: Hong Kong University Press.
22. Yang, R., & Wang, W. S. Y. (2018). Categorical perception of Chinese characters by simplified and traditional Chinese readers. *Reading & Writing*. 31(5), 1133–1154. https://doi.org/10.1007/s11145-018-9832-y.
23. The WHOQOL Group. (1993). Study protocol for the World Health Organization project to develop a quality of life assessment instrument (WHOQOL). *Quality of Life Research*. 2(2), 153–159. https://doi.org/10.1007/BF00435734.
24. Cha, E. S., Kim, K. H., & Erlen, J. A. (2007). Translation of scales in cross-cultural research: Issues and techniques. *Journal of Advanced Nursing*. 58(4), 386–395. https://doi.org/10.1111/j.1365-2648.2007.04242.x.
25. Varni, J. W., Burwinkle T. M., Seid, M., & Skarr, D. (2003). The PedsQL™ 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambulatory Pediatrics*. 3(6), 329–341. https://doi.org/10.1367/1539-4409(2003)003%3C0329:TPAAPP%3E2.0.CO;2.
26. Varni, J. W., Seid, M., & Rode, C. A. (1999). The PedsQL™: Measurement model for the pediatric quality of life. *Inventory Medical Care, 37*(2), 126–139. https://doi.org/10.1097/00005650-1999020 00–00003.
27. 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*(8), 1229–1233. https://doi.org/10.1007/s11136-010-9672-y.
28. Yeung, N. C., Lau, J. T., Yu, X. N., Chu, Y., Shing, M. M., Leung, T. F., et al. (2015). Psychometric properties of the Chinese version of the pediatric quality of life inventory 4.0 generic core scales among pediatric cancer patients. *Cancer Nursing*. 36(6), 463–473. https://doi.org/10.1097/NCC.0b013e31827028c8.
29. Currie, C., Elton, R. A., Todd, J., & Platt, S. (1997). Indicators of socioeconomic status for adolescents: the WHO health behaviour in school-aged children survey. *Health Education Research, 12*(3), 385–397. https://doi.org/10.1093/her/12.3.385.

30. Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) Family Affluence Scale. *Social Science & Medicine*. 66(6), 1429–1436. https://doi.org/10.1016/j.socscimed.2007.11.024.

31. Liu, Y., Wang, M., Villberg, J., Torsheim, T., Tynjälä, J., Lv, Y., et al. (2012). Reliability and validity of family affluence scale (FAS II) among adolescents in Beijing, China. *Child Indicators Research, 5*(2), 235–251. https://doi.org/10.1007/s12187-011-9131-5.

32. Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology & Psychiatry, 38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x.

33. Du, Y., Kou, J., & Coghill, D. (2008). The validity, reliability and normative scores of the parent, teacher and self-report versions of the strengths and difficulties questionnaire in china. *Child and Adolescent Psychiatry and Mental Health, 2*(1), 1–15. https://doi.org/10.1186/1753-2000-2-8.

34. Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry, 15*(1–2), 166–172. https://doi.org/10.1080/0954026021000046128.

35. Koskelainen, M., Sourander, A., & Kaljonen, A. (2000). The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *European Child & Adolescent Psychiatry, 9*(4), 277–284. https://doi.org/10.1007/s007870070031.

36. Muris, P., Meesters, C., & Berg, F. V. D. (2003). The Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry, 12*(1), 1–8. https://doi.org/10.1007/s00787-003-0298-2.

37. Lai, K. Y. C., Leung, P. W. L., Luk, E. S. L., & Wong, A. S. L. (2014). Use of the extended strengths and difficulties questionnaire (SDQ) to predict psychiatric caseness in Hong Kong. *Child Psychiatry & Human Development, 45*(6), 703–711. https://doi.org/10.1007/s10578-014-0439-5.

38. Liu, S. K., Chien, Y. L., Shang, C. Y., Lin, C. H., Liu, Y. C., & Gau, S. S. (2013). Psychometric properties of the chinese version of strength and difficulties questionnaire. *Comprehensive Psychiatry, 54*(6), 720–730. https://doi.org/10.1016/j.comppsych.2013.01.002.

39. Wong, G. K. C., Lam, S. W., Ngai, K., Wong, A., Poon, W. S., & Mok, V. (2013). Development of a short form of stroke-specific quality of life scale for patients after aneurysmal subarachnoid hemorrhage. *Journal of the Neurological Sciences*. 335(1–2), 204–209. https://doi.org/10.1016/j.jns.2013.09.033.

40. Rohenkohl, A. C., Bullinger, M., Pleil, A. M., Kriston, L., Quitmann, J. H., & the QoLISSY Study Group (2015). A Brief Version of the quality of life in short stature youth questionnaire–the QoLISSY-brief. *Child Indicators Research*, 9(4), 971–984. https://doi.org/10.1007/s12187-015-9350-2.

41. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.

42. Kline, P. (1993). *The handbook of psychological testing* (2nd edn.). London: Routledge.

43. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

44. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*. 6(1), 1–55. https://doi.org/10.1080/10705519909540118.

45. Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association. https://doi.org/10.1037/10694-000.

46. Byrne, B. M. (2001). Structural equation modeling with AMOS, EQS and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing, 1*(1), 55–86. https://doi.org/10.1207/S15327574IJT0101_4.

47. Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling A Multidisciplinary Journal*. 14(3), 464–504. https://doi.org/10.1080/10705510701301834.

48. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2n edn.). Hillsdale: Lawrence Erlbaum Associates. https://doi.org/10.1016/B978-0-12-179060-8.50012-8.

49. Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146.

50. Berra, S., Ravens-Sieberer, U., Erhart, M., Tebé, C., Bisegger, C., Duer, W., et al. (2007). Methods and representativeness of a European survey in children and adolescents: the KIDSCREEN study. *BMC Public Health, 7*(1), 1–12. https://doi.org/10.1186/1471-2458-7-182.

51. Haraldstad, K., Christophersen, K. A., Eide, H., Nativg, G. K., Helseth, S., & the KIDSCREEN Group Europed (2011). Health related quality of life in children and adolescents: Reliability and validity of the Norwegian version of KIDSCREEN-52 questionnaire, a cross sectional study. *International Journal of Nursing Studies, 48*(5), 573–581. https://doi.org/10.1016/j.ijnurstu.2010.10.001.

52. Nezu, S., Iwasaka, H., Saeki, K., Ishizuka, R., Goma, H., Okamoto, N., et al. (2015). Reliability and validity of the Japanese version of the KIDSCREEN-52 health-related quality of life questionnaire for children/adolescents and parents/proxies. *Environmental Health & Preventive Medicine, 20*(1), 44–52. https://doi.org/10.1007/s12199-014-0427-1.

53. Ravens-Sieberer, U., Erhart, M., Rajmil, L., Herdman, M., Auquier, P., Bruil, J., et al. (2010). Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. *Quality of Life Research, 19*(10), 1487–1500. https://doi.org/10.1007/s11136-010-9706-5.

54. Ravens-Sieberer, U., Wille, N., Badia, X., Bonsel, G., BurstrÓ§m, K., Cavrini, G., et al. (2010). Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Quality of Life Research, 19*(6), 887–897. https://doi.org/10.1007/s11136-010-9649-x.

55. Upton, P., Maddocks, A., Eiser, C., Barnes, P. M., & Williams, J. (2005). Development of a measure of the health-related quality of life of children in public care. *Child Care Health & Development, 31*(4), 409–415. https://doi.org/10.1111/j.1365-2214.2005.00520.x.

56. Parizi, A. S., Garmaroudi, G., Fazel, M., Omidvari, S., Azin, S. A., Montazeri, A., et al. (2014). Psychometric properties of KIDSCREEN health-related quality of life questionnaire in Iranian adolescents. *Quality of Life Research, 23*(7), 2133–2138. https://doi.org/10.1007/s11136-014-0655-2.

57. Taliep, N. (2010). *Evaluating the construct validity of the KIDSCREEN-52 quality of life questionnaire within a South African context utilising exploratory factor analysis: Initial validation*. Unpublished master's thesis, University of the Western Cape.

58. Tebé, C., Berra, S., Herdman, M., Aymerich, M., Alonso, J., & Rajmil, L. (2008). Reliability and validity of the Spanish version of the KIDSCREEN-52 for child and adolescent population. *Medicina Clinica, 130*(17), 650–654. https://doi.org/10.1157/13120999.

59. Tzavara, C., Tzonou, A., Zervas, I., Ravens-Sieberer, U., Dimitrakaki, C., & Tountas, Y. (2012). Reliability and validity of the KIDSCREEN-52 health-related quality of life questionnaire in a Greek adolescent population. *Annals of General Psychiatry, 13*(11), 3. https://doi.org/10.1186/1744-859X-11-3.

60. Baydur, H., Ergin, D., Gerceklioglu, G., & Eser, E. (2016). Reliability and validity study of the KIDSCREEN health-related quality of life questionnaire in a Turkish child and adolescent population. *Anatolian Journal of Psychiatry, 17*(6), 496–505. https://doi.org/10.5455/apd.214559.

61. Sepúlveda, P. R., Molina, G. T., Molina, C. R., Martínez, N. V., Gonzalez, A. E., George, L. M., et al. (2013). Validation of an instrument to measure health-related quality of life in Chilean children and adolescents. Revista Médica De Chile. 141(10), 1283–1292. https://doi.org/10.4067/S0034-98872013001000007.

62. Rigdon, E. E. (1995). A Necessary and Sufficient Identification Rule for Structural Models Estimated in Practice. *Multivariate Behavioral Research*, 30(3), 359–383. https://doi.org/10.1207/s15327906mbr3003_4.

63. Bagheri, Z., Jafari, P., Tashakor, E., Kouhpayeh, A., & Riazi, H. (2014). Assessing Whether Measurement Invariance of the KIDSCREEN-27 across Child-Parent Dyad Depends on the Child Gender: A Multiple Group Confirmatory Factor Analysis. *Global Journal of Health Science, 6*(5), 142–153. https://doi.org/10.5539/gjhs.v6n5p142.

64. Byrne, B. M. (2004). Testing for Multigroup Invariance Using AMOS Graphics: A Road Less Traveled. *Structural Equation Modeling A Multidisciplinary Journal*. 11(2), 272–300. https://doi.org/10.1207/s15328007sem1102_8.

65. Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187–212. https://doi.org/10.1177/0022022100031002003.

66. Hong, S. D., Yang, J. W., Jang, W. S., Byun, H., Lee, M. S., Kim, H. S., et al. (2007). The KIDSCREEN-52 quality of life measure for children and adolescents (KIDSCREEN-52-HRQOL): Reliability and validity of the Korean version. *Journal of Korean Medical Science, 22*(3), 446–452. https://doi.org/10.3346/jkms.2007.22.3.446.

67. Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Power, M., et al. (2007). The KIDSCREEN-52 quality of life measure for children and adolescents: Psychometric results from a cross-cultural survey in 13 European countries. *Value in Health*. 11(4), 645–658. https://doi.org/10.1007/s11136-007-9240-2.

68. Mechanic, D., & Hansell, S. (1987). Adolescent competence, psychological well-being, and self-assessed physical health. *Journal of Health & Social Behavior*. 28(4), 364–374. https://doi.org/10.2307/2136790.

69. Andersen, J. R., Natvig, G. K., Haraldstad, K., Skrede, T., Aadland, E., & Resaland, G. K. (2016). Psychometric properties of the Norwegian version of the kidscreen-27 questionnaire. *Health & Quality of Life Outcomes, 14*(1), 1–6. https://doi.org/10.1186/s12955-016-0460-4.