Validation of the Chinese Version of KIDSCREEN-10 Quality of Life Questionnaire: A Rasch Model Estimation

Zepeng Gong[1], Jia Xue[2], Ziqiang Han[3] and Yuhuan Li[4]*

[1] School of Public Affairs and Administration & Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Chengdu, China, [2] Factor-Inwentash Faculty of Social Work & Faculty of Information, University of Toronto, Toronto, ON, Canada, [3] School of Political Science and Public Administration, Shandong University, Qingdao, China, [4] School of Government, Central University of Finance and Economics, Beijing, China

The KIDSCREEN-10 was deemed as a cross-national instrument for measuring Health-Related Quality of Life (HRQoL). However, no empirical endeavor has explored its reliability and validity in the context of China. This study aims to translate and validate the Chinese version of the KIDSCREEN-10 questionnaire. The KIDSCREEN-10 was translated into Chinese (Mandarin) using a blindly bilingual forward–backward–forward technique. A cross-sectional survey, including 1,830 students aged from 8 to 18 years, was conducted in a county located in Gansu province, China. Psychometric properties were evaluated using the Rasch partial credit model, ANOVA, and the correlation analysis. Results indicated that the KIDSCREEN-10 performed good internal consistency, known-group validity, and concurrent validity, but there were still some deficiencies in psychometrics: first, disordered response categories were found between category 2 (seldom) and category 3 (sometimes); second, item 3 ("Have you felt sad?"), item 4 ("Have you felt lonely?"), and item 5 ("Have enough time for self?") demonstrated misfit to the Rasch model; third, items 3 and 4 exhibited differential item functioning. After collapsing the disordered response categories and removing the three misfit items, the seven-item questionnaire performed good psychometric properties. However, the seven-item version does not cover the psychological well-being dimension of HRQoL, and that may lead to inappropriate measures of HRQoL. Therefore, this paper suggested to use classical test theory to investigate the psychological properties of the KIDSCREEN-10.

Keywords: quality of life, KIDSCREEN-10, psychometric property, Rasch analysis, China

INTRODUCTION

Quality of life (QoL) is an important public health issue for the policy development (Phillips, 2006). QoL refers to the perception of subjective health and well-being of an individual, which is a multidimensional concept covering several dimensions, including social relationships, physical and psychological health (The World Health Organization, 1995). Assessing the QoL more scientifically is an essential step to improve QoL of people's, and researchers have invested lots of effort into developing appropriate assessment tools for the general public or specific groups. For

example, there are generic- and condition-specific instruments for evaluating QoL of children and adolescents. The former one is applicable to all population subgroups, whereas the latter one is useful to those with specific disability or illness (Fava et al., 2009; Davis et al., 2013; Bullinger et al., 2015). This study focused on the generic QoL instrument for children and adolescents.

Currently, the KIDSCREEN (The KIDSCREEN Group Europe, 2006), the Pediatric Quality of Life Inventory 4.0 (PedsQL 4.0) (Varni et al., 2006), the KINDL (Ravens-Sieberer and Bullinger, 1998), the Child Health Questionnaire (CHQ) (Raat et al., 2002), and Brief Multidimensional Students' Life Satisfaction Scale (BMSLSS) (Huebner, 1994) are well-known generic questionnaires that have been adopted to assess QoL in children and adolescents. These measures share some commonalities, but each one has its preference. The KIDSCREEN was designed to measure health-related quality of life (HRQoL), the PedsQL 4.0 included a wide definition of functioning, disability, and health (FDH), the KINDL, and the CHQ were appropriate to evaluate FDH with some HRQoL features, and the BMSLSS focused on life satisfaction (Seligson et al., 2003; Alamolhoda et al., 2021). To our knowledge, psychometric properties of the PedsQL 4.0, the KNIDL, the CHQ, and the BMSLSS have been validated in China (Ng et al., 2005; Lin et al., 2012, 2014; Ye et al., 2014), where has a population of 321 million children and adolescents under 20 years old (National Bureau of Statistics, 2011), whereas the KIDSCREEN has not been fully tested in this context.

KIDSCREEN instruments include self-report and proxy (parents) versions, and each of the two versions has three forms with 52, 27, and 10 items, respectively. The tools were initially and simultaneously developed in 13 European countries (The KIDSCREEN Group Europe, 2006). The KIDSCREEN-52 instrument with 52 items assesses 10 dimensions of HRQoL: physical well-being, psychological well-being, moods and emotions, self-perception, autonomy, parent relation and home life, financial resources, peers and social support, school environment, and bullying (Ravens-Sieberer et al., 2008; Zhu et al., 2019). The KIDSCREEN-27 is a short version of KIDSCREEN-52 using 27 items to measure five facets (i.e., physical well-being, psychological well-being, autonomy and parent relation, social support and peers, and school environment) merged from the 10 dimensions mentioned above (Ng et al., 2015). The KIDSCREEN-10 comprises 10 items derived from the 27-item version (The KIDSCREEN Group Europe, 2006). Evidence from prior studies indicated that KIDSCREEN-10 results in one global HRQoL score (Ravens-Sieberer et al., 2010; Haraldstad et al., 2011).

The KIDSCREEN instruments are deemed as cross-national HRQoL measures, and their psychometric properties have been studied in considerable research. Ravens-Sieberer et al. have conducted a cross-cultural survey in 13 European countries to assess the reliability and validity of KIDSCREEN indexes, and the result showed that all three versions of KIDSCREEN were reliable and valid (Robitail et al., 2007; Ravens-Sieberer et al., 2008, Ravens-Sieberer et al., 2010). Moreover, such supportive result for KIDSCREEN-52 has been found in investigations from China (Ng et al., 2015; Zhu et al., 2019), South Korea (Hong

et al., 2007), South Africa (Taliep and Florence, 2012), Serbia (Stevanovic et al., 2013), Iran (Parizi et al., 2014), Japan (Nezu et al., 2015), Turkey (Baydur et al., 2016), and Colombia (Jaimes-Valencia et al., 2019, 52). In addition, studies from China (Ng et al., 2015), Serbia (Stevanovic et al., 2013), Turkey (Baydur et al., 2016), Norway (Andersen et al., 2016), Japan (Nezu et al., 2016), and Colombia (Vélez et al., 2016) have found a similar result for KIDSCREEN-27. Regarding KIDSCREEN-10, its validity and reliability can be supported in studies from Serbia (Stevanovic et al., 2013), Turkey (Baydur et al., 2016), Japan (Nezu et al., 2016), and Iran (Nik-Azin et al., 2014). Comparatively, a little evidence evaluated measurement properties of KIDSCREEN-10, especially no empirical endeavor has explored its reliability and validity in China. Given KIDSCREEN-10 was recommended for large epidemiological studies (The KIDSCREEN Group Europe, 2006), it is necessary to translate and test the applicability of KIDSCREEN-10 in more countries.

The current study aims to validate the cross-cultural adaption of the Mandarin Chinese self-report questionnaire of KIDSCREEN-10 using the Rasch model. Although the KIDSCREEN-52 and the KIDSCREEN-27 have been tested in the context of China (Ng et al., 2015; Zhu et al., 2019), we cannot infer that the KIDSCREEN-10 originally developed them also has good psychometric properties. For example, the KIDSCREEN-52 and the KIDSCREEN-27 are multidimensional scales, whereas the KIDSCREEN-10 is deemed as an unidimensional measure (Ravens-Sieberer et al., 2014). Therefore, based on the results of Zhu et al. (2019) and Ng et al. (2015), we still do not know whether the KIDSCREEN-10 is unidimensional. Moreover, the Rasch model is an approach exploring the performance of each item rather than the total test score, as in the classical test theory (CTT) (da Rocha et al., 2013). The Rasch model provides a detailed analysis of how items work within scales (Tennant et al., 2004), and thus it has many potential advantages over CTT methods in evaluating self-reported health outcomes (Hays et al., 2000). Currently, it has been increasingly applied in the psychology and health fields (Rocha et al., 2012; Ng et al., 2015; Vélez et al., 2016). Previous studies have demonstrated that this approach is appropriate, and actually more accurate to discover the psychometric properties of KIDSCREEN-10 (Erhart et al., 2009; Ravens-Sieberer et al., 2014). Therefore, the Rasch model is adopted to examine the properties of the KIDSCREEN-10 instrument for measuring the HRQoL among Chinese children and adolescents.

METHODS

Sampling and Participants

We conducted a survey in students of all grades of middle, high, vocational schools and in Grade four to Grade six students of primary schools in a county of Gansu province, China. All participants completed the online questionnaires in the computer room of their schools with the help of research assistants and teachers. In total, 2,155 students participated our survey. After dropped those with missing value or aged above 18 years, 1,830 eligible respondents were obtained. Of the respondents, 50.98% were girls, and 38.31% were 8–12 years old and 61.69% aged from

TABLE 1 | Sample characteristics.

	Frequency	Percentage
Gender		
Female	933	50.98
Male	897	49.02
Ethnicity		
Minorities	1,170	63.93
Majority	660	36.07
Age		
8–12 years	701	38.31
13–18 years	1,129	61.69
School type		
Primary school	721	39.40
Middle school	607	33.17
High school	352	19.23
Vocational school[+]	150	8.20
Socioeconomic status		
Low	16	0.87
Medium to low	149	8.14
Medium	1,043	56.99
High to medium	565	30.87
High	57	3.11
Health status		
Very bad	7	0.38
Bad	12	0.66
Moderate	164	8.96
Good	831	45.41
Very good	816	44.59
Academic performance		
Very bad	71	3.88
Bad	233	12.73
Moderate	694	37.92
Good	674	36.83
Very good	158	8.63

[+]vocational school is equivalent to high school but focusing on working skills training.

12 to 18 years (**Table 1**). Students from primary, middle, high, and vocational schools accounted for 39.40, 33.17, 19.23, and 8.20%, respectively. Regarding ethnic minority group differences, a total of 660 students were Han (the national majority in China), 521 were Yugur, 568 were Tibetan, and 81 were other ethnic minorities, including the Uygur, Mongolian, Hui, and others.

Instruments
KIDSCREEN-10
As mentioned above, KIDSCREEN-10 is a 10-item self-report questionnaire (**Table 2**). The same five-point Likert scale measures each item. In the current study, participants were asked: "In the last week, how often do you experience the following items?" The answers to each item were: never (score = 1), seldom (score = 2), sometimes (score = 3), often (score = 4), and always (score = 5). The total score of all the 10 items was calculated to assess HRQoL. A higher total score indicates a

TABLE 2 | Items characteristics.

Items	Mean	SD	Floor (%)	Ceiling (%)
Item 1: felt fit and well	3.82	1.29	10.16	38.85
Item 2: felt full of energy	3.68	1.31	10.71	33.83
Item 3: felt sad	2.22	1.14	30.93	6.23
Item 4: felt lonely	2.09	1.18	39.18	6.39
Item 5: had enough time for self	2.89	1.30	16.89	15.08
Item 6: did things in free time	3.58	1.26	9.18	29.02
Item 7: treated fairly by parent(s)	3.94	1.29	9.07	46.17
Item 8: had fun with friends	4.11	1.15	6.56	48.03
Item 9: got on well at school	3.68	1.12	5.74	27.16
Item 10: pay attention	3.71	1.15	6.45	28.03

higher level of HRQoL. In order to ensure that Chinese students easily understand each item, the Mandarin Chinese version of KIDSCREEN-10 was translated from the English version using a blindly bilingual forward–backward–forward technique (**Supplementary Table 1**) (Brislin, 1970). First, two graduate students with good English and Chinese language skills translated the English version of KIDSCREEN-10 into Chinese (Mandarin) independently, and then they discussed their translation results with the supervisor and reached a consensus. Second, the Chinese version was given to another two graduate students for back-translation independently, and they also discussed and reached a consensus. Finally, a bilingual expert read and checked all the translation documents and confirmed the final version. The Mandarin KIDSCREEN-10 questionnaire requires ~3–5 min to complete.

Brief Multidimensional Students' Life Satisfaction Scale
The BMSLSS was designed to assess the satisfaction of five life domains (i.e., family, friends, school, self, and environment) of students aged from 8 to 18 years (Seligson et al., 2003). In the present study, participants were asked to evaluate how satisfied with the five areas mentioned above. The responses to each area were ranged from "extremely dissatisfied (score = 1)" to "extremely satisfied (score = 5)." The participant with a higher total score calculated by summing the scores of five areas indicated he/she had a higher life satisfaction. The BMSLSS has been proved to have good psychometric properties in the context of China in prior studies (Ye et al., 2014; Tian et al., 2015). The Cronbach's alpha of BMSLSS in this study was 0.95.

Other Variables
Socio-demographics, including gender, ethnicity, and age, were investigated. Moreover, students reported that they perceived socioeconomic status, health status, and academic performance. Socioeconomic status was measured from the question, "Compared with other families in your living region, how do you think of your family's socioeconomic level?" Of the respondents, more than half of them perceived the economic status of their families were medium, and nearly one-third perceived that their status was above medium

(**Table 1**). Regarding health status and academic performance, two questions, "how do you think of your health/academic performance?" were asked. Answers for these two questions were: very bad, bad, moderate, good, and very good. Most of the participants perceived their health as good or very good, and perceived their academic performance as moderate or above.

Statistical Analysis

Descriptive Analysis

First, the frequency and percentage of socio-demographic variables were reported. Then, the mean score (SD) and distribution of the 10 items of KIDSCREEN-10 were examined. A floor or ceiling effects were considered significant if more than 20% of the participants responded to the item with the answer of never or always (Holmes and Shea, 1997).

Rasch Analysis

The Rasch partial credit model for ordered response categories was adopted to evaluate the measurement properties of KIDSCREEN-10 by using the WINSTEPS 4.0.1 software (Beaverton, Oregon: Winsteps.com). In this study, we evaluated six aspects of the scale: response format, item fit, person reliability, unidimensionality, item difficulty, and differential item functioning (DIF). First, we tested the appropriateness of the response format of KIDSCREEN-10 by checking the order of step difficulty. Step difficulty refers to the threshold between adjacent response categories of an item (Linacre, 1999). The response format (never = 1, seldom = 2, sometimes = 3, often = 4, and always = 5) was considered to be appropriate when the step difficulties were properly ordered (i.e., no disorder). If any disorder was detected between response categories, we merged the disordered categories and reexamined the step difficulty of response categories to obtain an appropriate response format. Second, infit mean square (MNSQ) and outfit MNSQ were used to evaluate item fit (i.e., data-model fit). The ideal value of infit and outfit MNSQ is 1, and the acceptable range of MNSQ value was set as 0.6 to 1.4 (Wong et al., 2011; Jervaeus et al., 2013). If the MNSQ value of any item exhibited unacceptable, we deleted the item and reconducted the analysis. Third, person-separation reliability was computed to estimate the internal consistency of KIDSCREEN-10. A reliability value >0.7 was considered to be adequate (Vélez et al., 2016). Fourth, unidimensionality means that all items in an instrument measure the same latent trait. To examine the KIDSCREEN-10 is unidimensionality, principal component analysis of residuals was conducted. The assumption of unidimensionality was considered to be acceptable if the eigenvalues of the first contrast was <2 (Wu et al., 2016). Finally, DIF reflects the potential bias of items understood by different groups. We should ensure that the response of participants to items based on the latent trait of interest regardless of other characteristics, such as gender or age (Baylor et al., 2011). Therefore, a well-performing item should not demonstrate DIF, i.e., one group of participants understand one item in the same way as another group of participants (Tennant et al., 2004). This study tested the DIF across gender (boys/girls), ethnicity (majority/minorities), and age (8–12 years/13–18 years). The DIF was detected when the absolute value of DIF contrast (the difference in the difficulty of an item between two groups) was >0.5 logits, and the p-value of Welch's test was lower than 0.05 (Bond et al., 2015).

Known-Group Validity

The known-group validity of KIDSCREEN-10 was assessed by comparing the discrepancy of HRQoL level between groups, such as different socioeconomic statuses, health statuses, and academic performance groups. One-way ANOVA was applied to examine the differences between groups. Eta square (η^2) was computed. We considered the effect size magnitudes of η^2 0.01 as small, 0.06 as moderate, and 0.14 as large (Zhu et al., 2019).

Concurrent Validity

The Pearson correlation coefficients between KIDSCREEN-10 scores and BMSLSS scores were calculated to evaluate the concurrent validity of KIDSCREEN-10. The coefficient intervals of 0.1–0.3, 0.3–0.5, and 0.5 or more were considered as low, moderate, and large, respectively (Wu et al., 2013; Zhu et al., 2019).

RESULTS

Item Characteristics

Characteristics of items are shown in **Table 2**. Item 8 had the largest mean value of 4.11, with a SD of 1.15. The mean value of other items ranged from 2.09 to 3.94. Additionally, floor effects were observed in item 3 and item 4, whereas other items demonstrated ceiling effects except item 5.
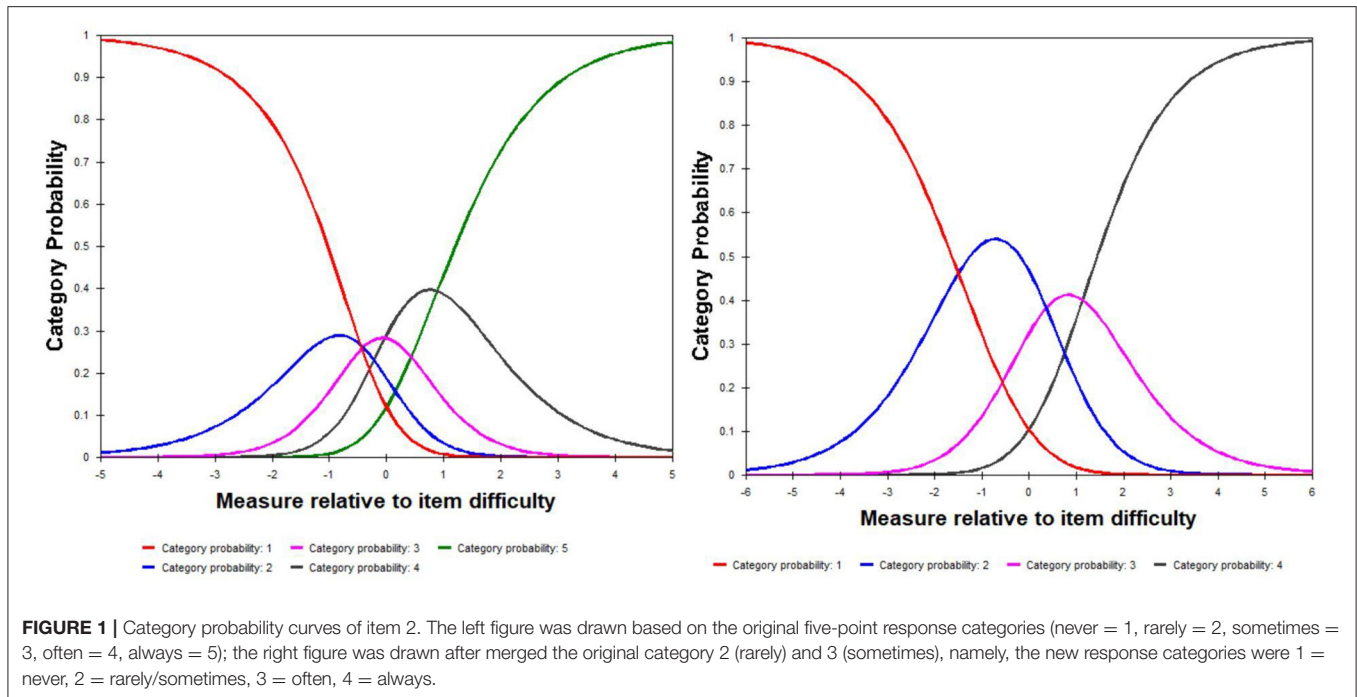
Response Format

Figure 1 shows the results of the step difficulty between adjacent response categories for item 2. There was category disorder in the left figure, which was drawn based on the original five-point response categories. After we collapsed the category 2 (seldom) and 3 (sometimes), then we reconducted Rasch analysis. The results from the right figure exhibited that no disorder presented in step difficulty.

Item Fit

The results of the fit analysis are shown in **Table 3**. In this study, three versions of the questionnaire were analyzed to fit the Rasch model. The version 1 was the original KIDSCREEN-10 with a five-point measurement scale. As we can see, infit (outfit) MNSQ value for item 3 and item 4 were 1.59 (1.72) and 1.58 (2.00), respectively. According to the results of the response format analysis, we should merge the response category 2 and category 3. Thus, the version 2 was the KIDSCREEN-10 using a four-point scale. Similarly, item 3 (outfit MNSQ = 1.62) and item 4 (outfit MNSQ = 1.91) did not fit the expectation of the Rasch model. Then, we removed these misfit items and reconducted analysis and found that item 5 was also a misfit item (infit MNSQ = 1.62, outfit MNSQ = 1.59). Accordingly, three items (i.e., item 3/4/5) were dropped from the KIDSCREEN-10. The version 3 was the questionnaire without any misfit item. It showed that the total fit (mean of infit MNSQ = 1.00, mean of outfit MNSQ = 1.01) of items was adequate. In addition, infit and outfit values of all items demonstrated acceptable fitness to the Rasch model.

FIGURE 1 | Category probability curves of item 2. The left figure was drawn based on the original five-point response categories (never = 1, rarely = 2, sometimes = 3, often = 4, always = 5); the right figure was drawn after merged the original category 2 (rarely) and 3 (sometimes), namely, the new response categories were 1 = never, 2 = rarely/sometimes, 3 = often, 4 = always.

TABLE 3 | Item difficulty and item fit statistics for KIDSCREEN before and after merged response categories 2 and 3.

	Version 1				Version 2				Version 3			
	Item difficulty	Model SE	Infit MNSQ	Outfit MNSQ	Item difficulty	Model SE	Infit MNSQ	Outfit MNSQ	Item difficulty	Model SE	Infit MNSQ	Outfit MNSQ
Item 1	−0.35	0.03	0.87	0.92	−0.48	0.03	0.83	0.88	0.09	0.04	1.00	1.04
Item 2	−0.23	0.03	0.77	0.79	−0.31	0.03	0.76	0.77	0.35	0.04	0.92	0.93
Item 3	1.28	0.03	1.59	1.72	1.67	0.04	1.58	1.62				
Item 4	1.37	0.03	1.58	2.00	1.79	0.04	1.62	1.91				
Item 5	0.48	0.02	1.00	1.02	0.63	0.03	1.02	1.01				
Item 6	−0.22	0.03	0.87	0.86	−0.28	0.03	0.86	0.85	0.45	0.04	1.30	1.33
Item 7	−0.49	0.03	0.92	0.90	−0.68	0.03	0.90	0.89	−0.20	0.04	1.10	1.15
Item 8	−0.80	0.03	0.81	0.77	−1.06	0.04	0.79	0.75	−0.70	0.04	0.86	0.84
Item 9	−0.55	0.03	0.79	0.79	−0.67	0.04	0.77	0.77	−0.02	0.04	0.94	0.94
Item 10	−0.48	0.03	0.77	0.78	−0.60	0.04	0.76	0.77	0.05	0.04	0.88	0.87
Mean	0.00	0.03	1.00	1.05	0.00	0.04	0.99	1.02	0.00	0.04	1.00	1.01
SD	0.73	0.00	0.30	0.41	0.96	0.00	0.32	0.39	0.35	0.00	0.14	0.16

Version 1 is the KIDSCREEN-10 with 10 items measured by five-point scale; version 2 includes 10 items measured by four-point scale (after merged response categories 2 and 3); and version 3 includes seven items measured by four-point scale; MNSQ: mean square; standard error (SE).

Reliability, Unidimensionality, and Item Difficulty

The person separation reliability value for version 1, 2, and 3 were 0.82, 0.85, and 0.83, respectively. These results reflected good internal consistency for three versions of the KIDSCREEN questionnaire. Regarding the unidimensionality, eigenvalue of the first contrast was 3.13 in the version 1, 2.91 in the version 2, and 1.98 in the version 3. Thus, only the version 3 exhibited unidimensionality. Moreover, in both version 1 and version 2, item 8 ("Have you had fun with your friends?") was the easiest item, whereas item 4 ("Have you felt lonely?") was the most

difficult item (**Table 3**). In the version 3, the easiest and most difficult items were item 8 (difficulty value = −0.70) and item 6 (difficulty value = 0.45), respectively.

Differential Item Functioning

The results of DIF tests are shown in **Table 4**. All items did not demonstrate DIF when we compared boys with girls, as well as the ethnicity of the majority with minorities. However, DIF was observed when comparing participants aged 8–12 years with 13–18 years, namely, item 4 ("have you felt lonely?") exhibited DIF in version 1 (DIF contrast value = 0.56, $p < 0.001$) and version

	Version 1			Version 2			Version 3		
	Gender	Ethnicity	Age	Gender	Ethnicity	Age	Gender	Ethnicity	Age
Item 1	0.07	−0.03	−0.07	0.11	−0.10	−0.07	0.14	−0.12	0.18**
Item 2	0.17***	0.00	−0.07*	0.25***	0.06	−0.09	0.35***	0.12	0.15*
Item 3	−0.12*	0.07	0.46***	−0.17*	0.11	0.68***			
Item 4	−0.08	0.04	0.56***	−0.12	0.11	0.86***			
Item 5	0.11*	0.00	0.23***	0.14	0.00	0.30***			
Item 6	0.00	0.06	−0.21***	0.00	0.08	−0.22**	0.00	0.16	−0.03
Item 7	−0.12*	−0.07	−0.14**	−0.16*	−0.07	−0.15*	−0.23**	−0.07	0.06
Item 8	0.00	−0.12	−0.18***	0.00	−0.14	−0.20**	0.00	−0.16	0.00
Item 9	−0.09	0.00	−0.31***	−0.10	0.00	−0.32***	−0.17	0.03	−0.19*
Item 10	−0.06	0.00	−0.29***	−0.08	0.00	−0.36***	−0.17	0.00	−0.24**

*gender (boys vs. girls); ethnicity (majority vs. minorities); age (8–12 vs. 13–18); DIF contrast values were reported; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

2 (DIF contrast value = 0.86, $p < 0.001$), whereas item 3 ("have you felt sad?" in version 2 (DIF contrast value = 0.68, $p < 0.001$).

Known-Group Validity

Differences in the KIDSCREEN scores by socioeconomic status, health status, and academic performance are shown in **Table 5**. The scores of the three versions exhibited a significant small effect size among participants with different socioeconomic statuses (η^2 ranged from 0.007 to 0.017) or with different academic performance (η^2 ranged from 0.041 to 0.058). However, the effect size on health status was moderate in version 3 ($\eta^2 = 0.073$).

Concurrent Validity

The Pearson correlation coefficients between the version 1 and the BMSLSS, and between the version 2 and the BMSLSS, exhibited a low coefficient effect size (0.26 for version 1; 0.28 for version 2). However, the coefficient (0.37) was moderate for the correlation between the version 3 and the BMSLSS.

DISCUSSION

This was the first study to examine the psychometric properties of KIDSCREEN-10 in the context of Chinese society. First, results found that the internal consistency of the questionnaire was acceptable, which were consistent with prior studies on KIDSCREEN-10 (Erhart et al., 2009; Ravens-Sieberer et al., 2010; Stevanovic et al., 2013; Nezu et al., 2016). Second, known-group validity was verified. The known-group validity, also named as construct validity, reflects that a test can discriminate between two groups known to vary on the variables of interest (Langevin, 2009; Hendriks et al., 2017). In this study, the HRQoL level measured by the KIDSCREEN-10 questionnaire demonstrated significant differences among different socioeconomic statuses/health statuses/academic performance groups. Similar differences across socioeconomic status were also found in studies from Turkey (Baydur et al., 2016) and European countries (Erhart et al., 2009; Ravens-Sieberer et al., 2010). Moreover, KIDSCREEN-10 has validity in predicting HRQoL levels of children and adolescents.

Seven items (i.e., items 1, 2, 6, 7, 8, 9, and 10) exhibited ceiling effects. Although previous studies did not exhibit such effects, these studies examined the overall ceiling effect of the questionnaire rather than each item (Nik-Azin et al., 2014; Baydur et al., 2016). The ceiling effects indicate that most of the items were easy for the respondents, and thus the KIDSCREEN-10 cannot differentiate well the respondents with high degrees of HRQoL. However, these items do not demonstrate floor effects. This is meaningful from a public health and clinical perspective because it is more important to differentiate well between respondents with low-QoL as these are of risks for various health problems and represent the target population for interventions.

The original KIDSCREEN-10 questionnaire measured by a 5-point Likert scale (i.e., response categories include "never," "seldom," "sometimes," "usually," and "always") demonstrated disordered response categories. It is because response options with similar meaning may make it difficult for respondents to distinguish the differences of options when answering, and thereby result in disordered categories (Zhong et al., 2014). After collapsing the categories of "seldom" and "sometimes," the response categories ordered appropriately. Therefore, the KIDSCREEN-10 questionnaire was better to be measured by the four response categories in China.

Inconsistent with prior finding that all items of KIDSCREEN-10 in 15 European countries exhibited a good fit to the Rasch partial credit model (Erhart et al., 2009), this study found that item 3 ("Have you felt sad?") and item 4 ("Have you felt lonely?") demonstrated misfit to the model. No matter before or after merging the response categories, both infit and outfit MNSQ values of items 3 and 4 were higher than the acceptable cut-off value of 1.4 (Zhong et al., 2014). Items misfit indicated that the response of these items was inconsistent with the overall response pattern (Liu et al., 2016). This might indicate that Chinese school-aged children's cognition of mental health was different from other aspects (such as activities participation, peer relationships, school performance, etc.) measured by KIDSCREEN-10 since these two items measured the mental health of a child (i.e., depressive moods and emotions and

TABLE 5 | Known-group validity tests.

	Version 1		Version 2		Version 3	
	η^2	F-value	η^2	F-value	η^2	F-value
Socioeconomic status	0.007	3.25*	0.009	4.21**	0.017	7.79***
Health status	0.022	10.40***	0.037	17.72***	0.073	36.01***
Academic performance	0.041	19.43***	0.046	22.06***	0.058	27.92***

eta square (η^2); *p < 0.05, **p < 0.01, ***p < 0.001.

stressful feelings) (Ravens-Sieberer et al., 2010). Meanwhile, this result might explain why the eigenvalue of the first contrast of the original KIDSCREEN-10 was higher than 2 thus violating the assumption of unidimensionality. Moreover, after removing these two items and reconducting the Rasch analysis, we found item 5 ("Have enough time for self?") was also a misfit item. Finally, the seven-item questionnaire assessed by a four-point Likert scale fitted the model well. It should be pointed out that, unlike deleting items 3 and 4 would delete a dimension (i.e., mental health) of the questionnaire, although item 5 was removed as well, the dimensions of the seven-item version did not decrease further. Because the seven-item questionnaire includes item 6, which reflects the same dimension (i.e., autonomy) as item 5 (Ravens-Sieberer et al., 2008, Ravens-Sieberer et al., 2014; Nezu et al., 2016).

Although the previous Rasch analyses showed the KIDSCREEN-10 was valid and reliable (Erhart et al., 2009; Ravens-Sieberer et al., 2014), they did not perform the DIF analysis. In the current study, both item 3 and item 4 exhibited DIF when comparing respondents aged from 8 to 11 years and 12 to 20 years. It can be implied that the use of the KIDSCREEN-10 in China was difficult because these items (items 3 and 4) cannot measure the HRQoL level of children and adolescents independently of age. Therefore, this study suggested that, to avoid the questionable conclusion, the DIF analysis should be conducted when the Rasch method is employed.

In summary, in contrast with previous studies that found the original KIDSCREEN-10 had acceptable validity and reliability in European countries by using the Rasch measurement model (Erhart et al., 2009; Ravens-Sieberer et al., 2014), our results showed that the psychometric features of the original KIDSCREEN-10 were deficiencies. This may be due to cultural differences. This study was conducted in the Chinese Mainland. China has a collectivist culture, which is different from the individualist culture of European countries. The psychometric properties of the seven-item version measured with four response categories ("never," "seldom/sometimes," "usually," and "always") performed better than the original KIDSCREEN-10, but it should be noted that the seven-item version does not contain items (i.e., items 3 and 4) on psychological well-being (Robitail et al., 2007), and thus it is not appropriate to measure the general HRQoL using the seven-item questionnaire. Accordingly, instead of advocating the seven-item version, this paper suggests that the psychological properties of the KIDSCREEN-10 should be further tested by using CTT. Previous studies demonstrated that KIDSCREEN-10 provided a CTT reliable and valuable assessment of general HRQoL (Erhart et al., 2009; Stevanovic

et al., 2013; Nik-Azin et al., 2014; Baydur et al., 2016; Nezu et al., 2016). Moreover, the KIDSCREEN-52, the parent version of the KIDSCREEN-10, has been validated by CTT in China (Ng et al., 2015; Zhu et al., 2019); thus, the KIDSCREEN-10 may not permit a Rasch-based measurement of general HRQoL, but fulfill the requirements of CTT.

This study has some limitations. First, the values of infits and outfits calculated by using the WINSTEPS software may lead to a high type I error rate because the software only calculates unconditional outfit and infit statistics, the results may become unreliable for sample sizes above 250 (Müller, 2020). Second, our sample, investigated from an autonomous county for ethnic minorities where the population of minorities was more than the national majority, may result in selection bias. Third, the concurrent validity of KIDSCREEN was examined by using the Pearson correlation coefficient. This method may under- or overestimate the predictive power, since it cannot adjust for confounders. Moreover, this study cannot provide the test–retest reliability due to the cross-sectional investigation.

CONCLUSIONS

The Mandarin version of the KIDSCREEN-10 did not perform good psychometrical properties in China by using the Rasch analysis. The KIDSCREEN-10 demonstrated disordered response categories, item misfit, unidimensionality, and DIF. After adjusting the response categories and removing the three misfit items, the seven-item questionnaire measured by the four response categories perform good measurement characteristics. However, the seven-item version was not appropriate to measure general HRQOL because it did not contain items on psychological well-being of HRQoL. Therefore, instead of advocating the seven-item version, this paper suggested that the psychological properties of the KIDSCREEN-10 should be further tested by using CTT.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional review board of Sichuan University (K2019067). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

ZG and YL proposed and designed this study. ZH collected the data. ZG and ZH analyzed the data and wrote the first draft. YL and JX made a revision. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.647692/full#supplementary-material

REFERENCES

Alamolhoda, M., Farjami, M., Bagheri, Z., Ghanizadeh, A., and Jafari, P. (2021). Assessing whether child and parent reports of the KINDL questionnaire measure the same constructs of quality of life in children with attention-deficit hyperactivity disorder. *Health Qual. Life Outcomes.* 19:19. doi: 10.1186/s12955-020-01649-w

Andersen, J. R., Natvig, G. K., Haraldstad, K., Skrede, T., Aadland, E., and Resaland, G. K. (2016). Psychometric properties of the Norwegian version of the Kidscreen-27 questionnaire. *Health Qual. Life Outcomes.* 14:58. doi: 10.1186/s12955-016-0460-4

Baydur, H., Ergin, D., Gerçeklioglu, G., and Eser, E. (2016). Reliability and validity study of the KIDSCREEN Health-Related Quality of Life Questionnaire in a Turkish child/adolescent population. *Anadolu. Psikiyatri. Dergisi.* 17:496–505. doi: 10.5455/apd.214559

Baylor, C., Hula, W., Donovan, N. J., Doyle, P. J., Kendall, D., and Yorkston, K. (2011). An introduction to item response theory and Rasch models for speech-language pathologists. *Am. J. Speech. Lang. Pathol.* 20:243–259. doi: 10.1044/1058-0360(2011/10-0079)

Bond, T., Fox, C. M., and Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* New York, Routledge. doi: 10.4324/9781315814698

Brislin, R. W. (1970). Back-translation for cross-cultural research. *J. Cross. Cult. Psychol.* 1:185–216. doi: 10.1177/135910457000100301

Bullinger, M., Sommer, R., Pleil, A., Mauras, N., Ross, J., Newfield, R., et al. (2015). Evaluation of the American-English Quality of Life in Short Stature Youth (QoLISSY) questionnaire in the United States. *Health Qual. Life Outcomes.* 13:43. doi: 10.1186/s12955-015-0236-2

da Rocha, N. S., Chachamovich, E., de Almeida Fleck, M. P., and Tennant, A. (2013). An introduction to Rasch analysis for Psychiatric practice and research. *J. Psychiatr. Res.* 47:141–148. doi: 10.1016/j.jpsychires.2012.09.014

Davis, E., Mackinnon, A., Davern, M., Boyd, R., Bohanna, I., Waters, E., et al. (2013). Description and psychometric properties of the CP QOL-Teen: a quality of life questionnaire for adolescents with cerebral palsy. *Res. Dev. Disabil.* 34:344–352. doi: 10.1016/j.ridd.2012.08.018

Erhart, M., Ottova, V., Gaspar, T., Jericek, H., Schnohr, C., Alikasifoglu, M., et al. (2009). Measuring mental health and well-being of school-children in 15 European countries using the KIDSCREEN-10 Index. *Int. J. Public Health.* 54:160–166. doi: 10.1007/s00038-009-5407-7

Fava, L., Muehlan, H., and Bullinger, M. (2009). Linking the DISABKIDS modules for health-related quality of life assessment with the International Classification of Functioning, Disability and Health (ICF). *Disabil. Rehabil.* 31:1943–1954. doi: 10.1080/09638280902874188

Haraldstad, K., Christophersen, K.-A., Eide, H., Nativg, G. K., and Helseth, S. (2011). Predictors of health-related quality of life in a sample of children and adolescents: a school survey. *J. Clin. Nurs.* 20:3048–3056. doi: 10.1111/j.1365-2702.2010.03693.x

Hays, R. D., Morales, L. S., and Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Med. Care.* 38:II28–II42. doi: 10.1097/00005650-200009002-00007

Hendriks, A. A. J., Smith, S. C., Chrysanthaki, T., and Black, N. (2017). Reliability and validity of a self-administration version of DEMQOL-Proxy. *Int. J. Geriatr. Psychiatry.* 32:734–741. doi: 10.1002/gps.4515

Holmes, W. C., and Shea, J. A. (1997). Performance of a new, HIV/AIDS-targeted quality of life (HAT-QoL) instrument in asymptomatic seropositive individuals. *Qual. Life Res.* 6:561–571. doi: 10.1023/A:1018464200708

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. *J. Korean Med. Sci.* 22:446–452. doi: 10.3346/jkms.2007.22.3.446

Huebner, E. S. (1994). Preliminary development and validation of a multidimensional life satisfaction scale for children. *Psychol. Assess.* 6:149–158. doi: 10.1037/1040-3590.6.2.149

Jaimes-Valencia, M. L., Perpiñá-Galvañ, J., Cabañero-Martínez, M. J., Cabrero-García, J., and Richart-Martínez, M. (2019). Adjusted linguistic validation and psychometric properties of the Colombian version of KIDSCREEN-52. *J. Child Health Care.* 23:20–34. doi: 10.1177/1367493518777291

Jervaeus, A., Kottorp, A., and Wettergren, L. (2013). Psychometric properties of KIDSCREEN-27 among childhood cancer survivors and age matched peers: a Rasch analysis. *Health Qual. Life Outcomes.* 11:96. doi: 10.1186/1477-7525-11-96

Langevin, M. (2009). The Peer Attitudes Toward Children who Stutter scale: reliability, known groups validity, and negativity of elementary school-age children's attitudes. *J. Fluency Disord.* 34:72–86. doi: 10.1016/j.jfludis.2009.05.001

Lin, C.-Y., Luh, W.-M., Cheng, C.-P., Yang, A.-L., and Ma, H.-I. (2014). Evaluating the wording effect and psychometric properties of the kid-KINDL. *Eur. J. Psychol. Assess.* 30:100–109. doi: 10.1027/1015-5759/a000175

Lin, C.-Y., Luh, W.-M., Yang, A.-L., Su, C.-T., Wang, J.-D., and Ma, H.-I. (2012). 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.* 21:177–182. doi: 10.1007/s11136-011-9928-1

Linacre, J. M. (1999). Investigating rating scale category utility. *J. Outcome Meas.* 3:103–122.

Liu, Y., Li, T., An, J., Zeng, W., and Xiao, S. (2016). Rasch analysis holds no brief for the use of the Dermatology Life Quality Index (DLQI) in Chinese neurodermatitis patients. *Health Qual. Life Outcomes.* 14:17. doi: 10.1186/s12955-016-0419-5

Müller, M. (2020). Item fit statistics for Rasch analysis: can we trust them? *J. Stat. Distrib. Appl.* 7:5. doi: 10.1186/s40488-020-00108-7

National Bureau of Statistics (2011). *Summary data of the Sixth National Census.* Available online at: http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm (accessed June 2, 2020).

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. *Environ. Health Prev. Med.* 20:44–52. doi: 10.1007/s12199-014-0427-1

Nezu, S., Iwasaka, H., Saeki, K., Obayashi, K., Ishizuka, R., Goma, H., et al. (2016). Reliability and validity of Japanese versions of KIDSCREEN-27 and KIDSCREEN-10 questionnaires. *Environ. Health Prev. Med.* 21:154–163. doi: 10.1007/s12199-016-0510-x

Ng, J. Y. Y., Burnett, A., Ha, A. S., and Sum, K. W. (2015). Psychometric properties of the Chinese (Cantonese) versions of the KIDSCREEN health-related quality of life questionnaire. *Qual. Life Res.* 24:2415–2421. doi: 10.1007/s11136-015-0973-z

Ng, J. Y. Y., Landgraf, J. M., Chiu, C. S. W., Cheng, N. L., and Cheung, Y. F. (2005). Preliminary evidence on the measurement properties of the Chinese version of the child health questionnaire, parent form (CHQ-PF50) and child form (CHQ-CF87). *Qual. Life Res.* 14:1775–1781. doi: 10.1007/s11136-005-1005-1

Nik-Azin, A., Shairi, M. R., Naeinian, M. R., and Sadeghpour, A. (2014). The health-related quality of life index KIDSCREEN-10: confirmatory factor analysis, convergent validity and reliability in a sample of iranian students. *Child. Ind. Res.* 7:407–420. doi: 10.1007/s12187-013-9216-4

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. *Qual. Life Res.* 23:2133–2138. doi: 10.1007/s11136-014-0655-2

Phillips, D. (2006). *Quality of Life: Concept, Policy and Practice.* London: Routledge. doi: 10.4324/9780203356630

Raat, H., Bonsel, G. J., Essink-Bot, M.-L., Landgraf, J. M., and Gemke, R. J. B. J. (2002). Reliability and validity of comprehensive health status measures in children: The Child Health Questionnaire in relation to the Health Utilities Index. *J. Clin. Epidemiol.* 55:67–76. doi: 10.1016/S0895-4356(01)00411-5

Ravens-Sieberer, U., and Bullinger, M. (1998). Assessing health-related quality of life in chronically ill children with the German KINDL: first psychometric and content analytical results. *Qual. Life Res.* 7:399–407. doi: 10.1023/A:1008853819715

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. *Qual. Life Res.* 19:1487–1500. doi: 10.1007/s11136-010-9706-5

Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Power, M., et al. (2008). The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. *Value Health.* 11:645–658. doi: 10.1111/j.1524-4733.2007.00291.x

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. *Qual. Life Res.* 23:791–803. doi: 10.1007/s11136-013-0428-3

Robitail, S., Ravens-Sieberer, U., Simeoni, M.-C., Rajmil, L., Bruil, J., Power, M., et al. (2007). Testing the structural and cross-cultural validity of the KIDSCREEN-27 quality of life questionnaire. *Qual. Life Res.* 16:1335–1345. doi: 10.1007/s11136-007-9241-1

Rocha, N. S., Power, M. J., Bushnell, D. M., and Fleck, M. P. (2012). Cross-cultural evaluation of the WHOQOL-BREF domains in primary care depressed patients using Rasch analysis. *Med. Decis. Making.* 32:41–55. doi: 10.1177/0272989X11415112

Seligson, J. L., Huebner, E. S., and Valois, R. F. (2003). Preliminary validation of the brief multidimensional students' life satisfaction scale (BMSLSS). *Soc. Indic. Res.* 61:121–145. doi: 10.1023/A:1021326822957

Stevanovic, D., Tadic, I., Novakovic, T., Kisic-Tepavcevic, D., and Ravens-Sieberer, U. (2013). Evaluating the Serbian version of the KIDSCREEN quality-of-life questionnaires: reliability, validity, and agreement between children's and parents' ratings. *Qual. Life Res.* 22:1729–1737. doi: 10.1007/s11136-012-0286-4

Taliep, N., and Florence, M. (2012). Evaluating the construct validity of the Kidscreen-52 quality of life questionnaire within a south African context. *S. Afr. J. Psychol.* 42:255–269. doi: 10.1177/008124631204200212

Tennant, A., McKenna, S. P., and Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value Health.* 7:S22–S26. doi: 10.1111/j.1524-4733.2004.7s106.x

The KIDSCREEN Group Europe (2006). *The KIDSCREEN Questionnaires: Quality of Life Questionnaires for Children and Adolescents.* Lengerich: PABST SCIENCE PUBLISHERS. Available online at: https://www.kidscreen.org (accessed November 18, 2019).

The World Health Organization (1995). The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. *Soc. Sci. Med.* 41:1403–1409. doi: 10.1016/0277-9536(95)00112-K

Tian, L., Zhang, J., and Huebner, E. S. (2015). Preliminary Validation of the brief multidimensional students' life satisfaction scale (BMSLSS) among Chinese elementary school students. *Child. Ind. Res.* 8:907–923. doi: 10.1007/s12187-014-9295-x

Varni, J. W., Burwinkle, T. M., and Seid, M. (2006). The PedsQLTM 4.0 as a school population health measure: feasibility, reliability, and validity. *Qual. Life Res.* 15:203–215. doi: 10.1007/s11136-005-1388-z

Vélez, C.-M., Lugo-Agudelo, L.-H., Hernández-Herrera, G.-N., and García-García, H.-I. (2016). Colombian Rasch validation of KIDSCREEN-27 quality of life questionnaire. *Health Qual. Life Outcomes.* 14:67. doi: 10.1186/s12955-016-0472-0

Wong, H. M., McGrath, C. P. J., and King, N. M. (2011). Rasch validation of the early childhood oral health impact scale. *Community Dent. Oral Epidemiol.* 39:449–457. doi: 10.1111/j.1600-0528.2011.00614.x

Wu, H., Li, H., and Gao, Q. (2013). Psychometric properties of the Chinese version of the pediatric quality of life inventory 4.0 Generic core scales among children with short stature. *Health Qual. Life Outcomes.* 11:87. doi: 10.1186/1477-7525-11-87

Wu, T.-Y., Yu, W.-H., Huang, C.-Y., Hou, W.-H., and Hsieh, C.-L. (2016). Rasch analysis of the general self-efficacy scale in workers with traumatic limb injuries. *J. Occup. Rehabil.* 26:332–339. doi: 10.1007/s10926-015-9617-y

Ye, M., Li, L., Li, Y., Shen, R., Wen, S., and Zhang, J. (2014). Life satisfaction of adolescents in Hunan, China: reliability and validity of chinese brief multidimensional students' life satisfaction scale (BMSLSS). *Soc. Indic. Res.* 118:515–522. doi: 10.1007/s11205-013-0438-0

Zhong, Q., Gelaye, B., Fann, J. R., Sanchez, S. E., and Williams, M. A. (2014). Cross-cultural validity of the Spanish version of PHQ-9 among pregnant Peruvian women: a Rasch item response theory analysis. *J. Affect. Disord.* 158:148–153. doi: 10.1016/j.jad.2014.02.012

Zhu, Y., Li, J., Hu, S., Li, X., Wu, D., and Teng, S. (2019). Psychometric properties of the Mandarin Chinese version of the KIDSCREEN-52 health-related quality of life questionnaire in adolescents: a cross-sectional study. *Qual. Life Res.* 28:1669–1683. doi: 10.1007/s11136-019-02158-0