



Psychometric evaluation of the Chinese version of the Child Health Utility 9D (CHU9D-CHN): a school-based study in China

Peirong Yang^{1,2} · Gang Chen³ · Peng Wang⁴ · Kejian Zhang² · Feng Deng² · Haifeng Yang² · Guihua Zhuang¹

Accepted: 25 April 2018 / Published online: 5 May 2018
© Springer International Publishing AG, part of Springer Nature 2018

Abstract

Purpose The Child Health Utility 9D (CHU9D), a new generic preference-based health-related quality of life (HRQoL) instrument, was developed specifically for the application in cost-effectiveness analyses of treatments and interventions for children and adolescents. The main objective of this study was to examine the psychometric property of the Chinese version of CHU9D (CHU9D-CHN) in a large school-based sample in China.

Methods Data were collected using a multi-stage sampling method from third-to-ninth-grade students in Shaanxi Province, China. Participants self-completed a hard-copy questionnaire including the CHU9D-CHN instrument, the Pediatric Quality of Life Inventory™ 4.0 Generic Core Scales (PedsQL), information on socio-demographic characteristics and self-reported health status. The psychometric properties of the CHU9D-CHN, including the internal consistency, 2-week test–retest reliability, convergent and known-groups validity were studied.

Results A total of 1912 students participated in the survey. The CHU9D-CHN internal consistency and test–retest reliability were good to excellent with a Cronbach’s alpha of 0.77 and an intra-class correlation coefficient of 0.65, respectively. The CHU9D utility scores moderately correlated with the PedsQL total scores ($r = .57, P < .001$), demonstrating good convergent validity. Difference of the CHU9D utility scores among the different participants with levels of self-reported general health, health services utilisation and left-behind status demonstrated good construct validity.

Conclusion The findings demonstrated adequate psychometric performance for the CHU9D-CHN. The CHU9D-CHN was a satisfactory, reliable and valid instrument to measure and value HRQoL for children and adolescents in China.

Keywords Health-related quality of life · Utility · CHU9D · Children · Adolescents · China

Introduction

Adolescence represents a critical development stage associated with more biological, psychological and social role changes than any other stage of life except infancy [1, 2]. During adolescence, individuals become increasingly responsible for their own health and health care; in the meantime, several health risk behaviours (e.g. alcohol use, cigarette smoking) start to become prevalent. Adolescence is therefore an important life phase where the introduction of targeted educational and preventative efforts has the potential to impact positively upon both short- and long-term health status and health-related quality of life outcomes (HRQoL) [3]. With recent revision in the family planning policy in China, it is expected that the number of children and adolescents will increase significantly, as well as the corresponding increasing demand for health services. There is a crucial and urgent need to understand children

✉ Gang Chen
gang.chen@monash.edu

✉ Guihua Zhuang
zhuanggh@mail.xjtu.edu.cn

¹ Department of Epidemiology and Biostatistics, School of Public Health, Xi’an Jiaotong University Health Science Center, No. 76 Yanta Western Road, Xi’an 710061, Shaanxi, China

² Baoji Center for Disease Control and Prevention, Baoji, Shaanxi, China

³ Centre for Health Economics, Monash University, 15 Innovation Walk, Clayton Campus, Melbourne, VIC 3800, Australia

⁴ College of Foreign Language, Baoji University of Arts and Sciences, Baoji, China

and adolescents' subjective valuation of their HRQoL. Such information is an essential prerequisite for the planning and prioritising of health-related interventions to improve adolescent health and also reflecting adolescent preferences.

Cost-utility analysis (CUA) is one type of economic valuation methods that has been commonly adopted to inform decision-makers about the relative benefits of alternative uses of resources in both clinical practice and public health settings [4, 5]. For an economic evaluation to allow for the comparability of outcomes emanating from interventions relating to disparate health conditions, the effectiveness of an intervention is usually quantified in terms of quality-adjusted life-years (QALYs). In recent years, generic preference-based instruments have become the most popular tools for the estimation of QALYs within CUA. A recent review by Chen and Ratcliffe [6] identified nine generic preference-based instruments available internationally that have been used in paediatric populations, among which the Child Health Utility 9D (CHU9D) is the only instrument that was developed from its inception exclusively for application with young people [7, 8].

The original CHU9D questionnaire was developed in the United Kingdom in English. Several recent studies have demonstrated the practicality, face and construct validity of the CHU9D in children and adolescents aged 7–17 years old from English-speaking countries [9–11]. There is an increasing interest internationally in the application of the CHU9D with adolescents. The instrument is currently being applied in a number of research programs, such as the economic evaluation of treatment programs for type 1 diabetes [12], mental health [13] and obesity prevention [14]. Recently, a Chinese version CHU9D (CHU9D-CHN) was developed and a pilot study was conducted in Nanjing City, eastern China suggesting that the CHU9D-CHN can discriminate self-reported health status of students in China [15]. The pilot study demonstrated the potential for using CHU9D-CHN in China. An official version of CHU9D-CHN questionnaire was translated after the pilot study by Oxford Outcomes Ltd, a research agency specialising in the translation and linguistic validation of patient-reported outcome measures in the United Kingdom. The translation and linguistic validation process were designed in line with the recommendations of the ISPOR Task Force for translation and cultural adaptation of patient-reported outcome measures [16], which include dual forward translation, reconciliation, dual back translation including back translation review, followed by cognitive debriefing with participants from the relevant patient group. The translation also underwent proofreading and format check steps prior to finalisation.

This study aims to evaluate the psychometric properties for the official CHU9D-CHN instrument in a large and representative school-based sample in western China. The availability of the CHU9D-CHN will facilitate future CUA

to prioritise health-related interventions for children and adolescents in China.

Methods

Participants

The study employed a cross-sectional school-based survey design in China, in which the targeted participants were students in grades 3–9. According to Kendall's rough working principle, the sample size should be 5–10 times the desirable variables. In the study, the targeted sample size was 198 students for every grade, which was 20 times more than the number of items in the CHU9D-CHN. Considering the rate of loss (10%), the total required theoretical sample size was 1386 participants. The participants were recruited using a multi-stage sampling method. First, we randomly chose four counties from Shaanxi Province in China. Second, within each chosen county, one school each was randomly selected from primary schools and secondary schools, respectively. Finally, two classes were randomly chosen in each grade.

Procedure

The survey was conducted by the Baoji Center for Disease Control and Prevention (CDC). Researchers underwent thorough training sessions prior to the survey. All the students in the school on the day of the survey completed the questionnaire independently in the classroom. Researchers were available to assist the students in completing the questionnaires if they had any semantic or conceptual understanding issues. Taking each school as a sampling unit, a primary school was selected for the retest survey by using a random sampling method. A total of 232 students (over 10% of the full sample) filled in the questionnaire again 2 weeks post the initial survey.

Questionnaire

The key components of the questionnaire include two generic HRQoL instruments (the CHU9D-CHN and the Pediatric Quality of Life Inventory™ 4.0 Generic Core Scales [hereafter "PedsQL"]), self-reported general health status and health service utilisation and socio-demographics characteristics, e.g. gender, age and left-behind status.

The CHU9D consisted of nine dimensions, including worried, sad, pain, tired, annoyed, schoolwork, sleep, daily routine and ability to join in activities. Within each dimension, there are five different levels indicating increasing levels of severity. The CHU9D-CHN was administered as a component of a paper questionnaire, and participants were instructed to self-complete the CHU9D-CHN from the

perspective of their own health on the particular day. The CHU9D-CHN questionnaire was scored using the latest version of the scoring algorithm derived from adolescents in Australia [17] and China [18–20] using a combination of the best worst scaling (BWS) and time trade-off (TTO) method. To derive the Chinese-specific tariff, two surveys were conducted in Nanjing, China. The first BWS survey recruited 902 students from primary and high schools using a multi-stage random sampling method and was administered in a classroom setting, whilst the TTO survey adopted an interviewer-administrated conventional TTO task and was administered to a convenience sample of 38 undergraduate students. A latent class-modelling framework which accounts for the potential preference heterogeneity of respondents was used to analyse the BWS data. Since the BWS approach generates raw health state values on an ordinal scale only, it was then re-scaled onto the 0–1 death–full health QALY scale by combining the responses from a secondary TTO sub-study using a mapping approach.

The Chinese version of the PedsQL is a validated and widely used non-preference-based paediatric HRQoL instrument [21–24]. The PedsQL consists of 23 items which can be grouped into 4 dimensions: physical health (8 items), emotional functioning (5 items), social functioning (5 items) and school functioning (5 items). The latter three dimensions can be used to further generate a psychosocial health summary score. A 5-point response scale is used in the report (0 = never a problem; 1 = almost never a problem; 2 = sometimes a problem; 3 = often a problem; 4 = almost always a problem). Items are reverse-scored and linearly transformed to a 0–100 scale (0 = 100, 1 = 75, 2 = 50, 3 = 25, 4 = 0), with higher scores indicating better HRQoL [25]. Scores of the PedsQL are computed as the sum of the items divided by the number of items answered (this accounts for missing data). If more than 50% of the items in the scale are missing, the score is not computed.

Analysis

Descriptive statistics including means, median, standard deviations (SD) and the interquartile range (IQR) were estimated. The distributions of the HRQoL scores obtained from each instrument were tested for normality using the Shapiro–Wilk test. Because the null hypothesis that scores of both instruments were non-normally distributed was rejected, the Kruskal–Wallis test and Mann–Whitney *U* tests were used to compare the scores for sub-groups. The two-tailed $P < .05$ was considered to be statistically significant. Statistical analysis was performed in SPSS version 22.0 (Armonk, NY: IBM Corp). The following psychometric properties of the CHU9D-CHN were investigated.

Investigation 1: reliability

The reliability was addressed by calculating Cronbach's alpha for the CHU9D-CHN [26]. Given that the CHU9D items tap into the same overall construct (quality of life) but represent different domains, we set an alpha of 0.7 as a minimally acceptable level of internal consistency [27]. It was anticipated that the CHU9D should have acceptable to excellent internal consistency indicating the items are tapping into the same general construct. Test–retest reliability (which shows agreement between participant's scores at two time intervals) was further investigated. The test–retest reliability was indexed by the intra-class correlation coefficient (ICC). An ICC > 0.60 was considered satisfactory [28]. A 2-week test–retest reliability was investigated following previous literature investigating the retest reliability of the most widely used preference-based HRQoL instrument among adults, the EQ-5D, in which the same length of recall period is used as the CHU9D [29–31]. Reliability was also assessed by examining the percentage agreement and kappa coefficients (weighted and unweighted) [32]. A kappa coefficient < 0.2 indicates poor agreement, 0.21–0.40 indicates fair agreement, 0.41–0.60 indicates moderate agreement, 0.61–0.80 indicates substantial agreement and > 0.81 indicates almost perfect agreement [33].

Investigation 2: convergent validity

We anticipate that the items and utility score on the CHU9D should demonstrate moderate to high correlations with corresponding similar items and score on the PedsQL, indicating that they are tapping into the same broad construct. For example, among CHU9D-CHN dimensions, worried, sad and annoyed are expected to have higher correlations with emotional function in the PedsQL; activities dimension is expected to have higher correlation with physical function in PedsQL and schoolwork/homework is expected to have a higher correlation with school function in the PedsQL. The association between the CHU9D and the PedsQL were investigated using Spearman's correlation coefficient. Correlations less than 0.3 were considered weak, 0.3–0.6 moderate and > 0.6 strong [34]. It should also be noted that the two instruments have different recall time periods: “today” was used for the CHU9D-CHN, whilst “during the past one month” is used for the PedsQL. The distribution of PedsQL scores across the dimension levels of the CHU9D-CHN was also studied. We generated a dimension-level correlation matrix of the CHU9D-CHN and the PedsQL to look for correlations between dimensions, particularly those that were conceptually related.

Investigation 3: known-groups validity

It was hypothesised that both the CHU9D and the PedsQL discriminate well between participants in different self-reported general health (excellent, very good, good, fair or poor), frequency of self-reported doctor visits and left-behind status. It was anticipated that participants with better health status would have significantly higher HRQoL scores (i.e. CHU9D utility scores and the PedsQL total score) than those who had worse health status. Similarly, we hypothesised that those who more frequently visit doctors should be associated with poorer HRQoL. Left-behind children or adolescents refer to those who are under 18 years old and are left behind at their original residence for at least 6 months by one or both parents who migrate to another place to work [35]. In China, the number of left-behind children has increased dramatically during the past 20 years, especially in rural areas [36]. Evidence reveals that left-behind children suffer from severe psychological problems, such as psychological depression or anxiety [35, 37], as well as behavioural issues [38, 39]. It was anticipated that participants who were left-behind would have lower HRQoL score than their counter-part.

Results

A total of 1959 students who attended the school on the day of the survey were eligible to participate and all successfully completed the survey. There were 40 (2.04%) questionnaires with missing items in the CHU9D-CHN. In addition, there were another seven questionnaires with missing items in self-reported general health status, health service utilisation and socio-demographics characteristics. The following psychometric analyses focused on a subsample ($N = 1912$) with no missing values on key items mentioned above. There were no significant differences between the analysis sample and full sample in participants' characteristics presented in Table 1.

For the analysis sample, the mean age \pm SD of the respondents was 12 ± 2 years old, of which 44% were female, and 64% were left-behind children. In this school-based sample, the majority of participants (82%) self-reported their health to be good (21%), very good (33%) or excellent (28%) and 66% did not visit doctors in the past month. The mean \pm SD (median) of the CHU9D-CHN utility scores were 0.78 ± 0.19 (0.81) and 0.85 ± 0.14 (0.88), scored using the Australian and Chinese adolescent tariffs, respectively (Fig. 1). The mean \pm SD (median) of the PedsQL score was 81.91 ± 12.55 (82.95). CHU9D utilities and PedsQL scores were not normally distributed (all $P < .001$).

The responses to the CHU9D-CHN are presented in Table 2. Participants also generally reported themselves in good health according to the CHU9D-CHN descriptive system, with $N = 328$ (17%) reporting themselves at full health

Table 1 Descriptive characteristics of participants

Characteristics	Analysis sample $N = 1912$ (%)
Gender	
Boys	1066 (55.75)
Girls	846 (44.25)
Age (years)	
8–12	1089 (56.96)
13–17	823 (43.04)
Self-reported general health status	
Excellent	540 (28.24)
Very good	633 (33.11)
Good	407 (21.29)
Fair or poor	332 (17.36)
Frequency of doctor visits in the last month (times)	
0	1253 (65.53)
1–2	582 (30.44)
≥ 3	77 (4.03)
Left-behind	
Yes	1225 (64.07)
No	687 (35.93)
CHU9D utility score, Australian tariff	
Mean (SD)	0.78 (0.19)
Median (Q_L – Q_U)	0.81 (0.67–0.92)
CHU9D utility score, Chinese tariff	
Mean (SD)	0.85 (0.14)
Median (Q_L – Q_U)	0.88 (0.78–0.96)
PedsQL score	
Mean(SD)	81.91 (12.55)
Median (Q_L – Q_U)	82.95 (73.91–92.39)

(reflecting the highest or best level for all nine CHU9D-CHN dimensions). No participants reported themselves in the lowest or worst level of impairment for all nine CHU9D dimensions. The dimension of the CHU9D-CHN with the highest proportion of no impairment was daily routine (89%), followed by sleep (81%), sad (76%), annoyed (74%), pain (71%), able to join in activities (59%), worried (55%), schoolwork/homework (55%) and tired (53%). It can also be seen that, along with the increasing levels of severity in the CHU9D-CHN dimensions, in most cases the mean PedsQL scores decreased. The three exceptions observed were the 'sad', the 'tired' and the 'sleep' dimensions, in which there were increased PedsQL scores from 'a bit' to 'quite', 'quite' to 'very' and from 'many problems' to 'couldn't sleep', respectively. However, only a very small proportion (<5%) of responses represent these inconsistent groups.

Reliability

The Cronbach's alpha was 0.771 (the standardised Cronbach's alpha coefficient was 0.780), indicating an acceptable

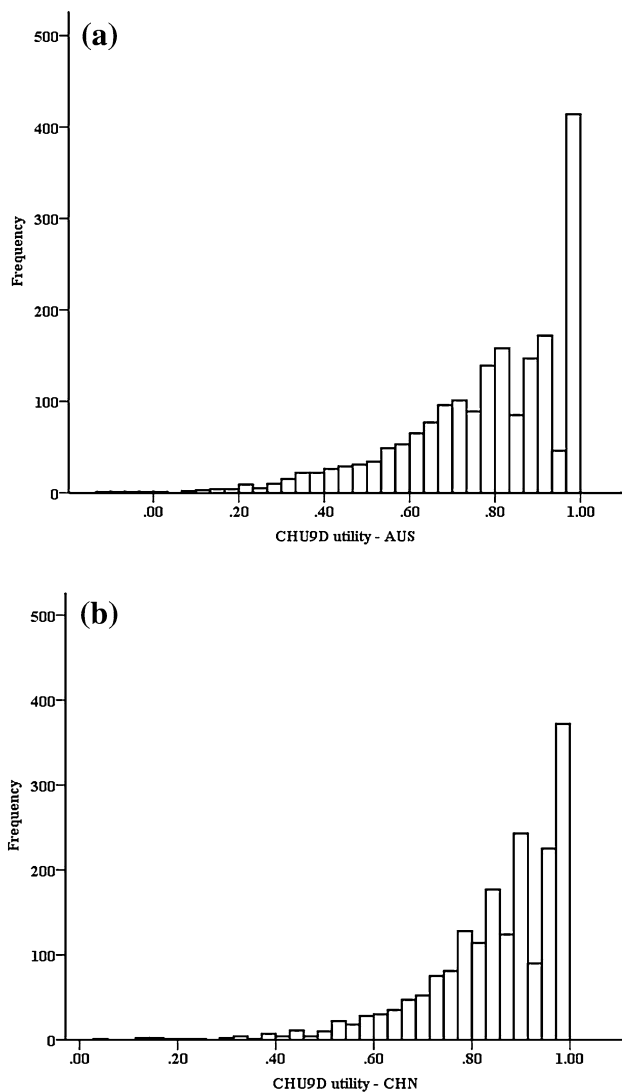


Fig. 1 Distribution of the CHU9D-CHN utility scores based on both Australian and Chinese tariffs

level of internal consistency for the CHU9D-CHN. A total of 232 students participated in the retest survey 2 weeks post the initial survey. The ICC was 0.653 ($P < .001$) for the overall CHU9D utility score, revealing good test–retest reliability. At dimension level, percentage agreement ranges from 49.57 (‘worried’ dimension) to 81.90% (‘daily routine’ dimensions), whilst weighted kappa coefficients ranged from poor to moderate (see Table 3 for details).

Convergent validity

Table 4 presents the correlations between CHU9D-CHN and PedsQL at both the overall scores and dimension levels. The CHU9D utility scores moderately correlated with

the PedsQL total scores ($r = .57$, $P < .001$). At the dimension level, the strongest degree of correlation was found between the ‘Emotional function’ dimension in the PedsQL and the ‘Worried’ dimensions on the CHU9D-CHN ($r = .37$; $P < .001$), which is classified as moderately correlated. A moderate degree of correlation was also found between similar dimensions for both instruments, ‘Schoolwork’ in the CHU9D-CHN and ‘School function’ in PedsQL ($r = .36$; $P < .001$), ‘Activities’ in the CHU9D-CHN and ‘Physical health’ in PedsQL ($r = .32$; $P < .001$). In addition, lower correlations were found between relatively dissimilar items of the CHU9D-CHN and dimensions of the PedsQL, showing evidence of discriminate validity. The lowest degree of correlations was found between the ‘Daily routine’ dimension in the CHU9D-CHN and the ‘School function’ dimension in the PedsQL ($r = .17$; $P < .01$). There was a consistently low agreement between all the items in the CHU9D-CHN and the ‘Social function’ dimension in the PedsQL ($r < .30$).

Known-groups validity

Table 5 summarises the differences of the CHU9D utilities and the PedsQL overall score between sub-groups based on individual characteristics. As expected, poor self-reported health status was significantly associated with lower CHU9D utilities ($P < .05$). An expected trend has also been found for health services utilisation in that those who used more health services were associated with lower CHU9D utilities ($P < .05$). Finally, the CHU9D utility scores of participants who were left-behind children were lower than their counterpart ($P < .05$).

Results on other demographic characteristics found that firstly, girls reported statistical and significant higher CHU9D utilities than boys ($P < .05$), and secondly, there was a decreasing trend for CHU9D utilities along with the increasing age ($P < .05$). Conclusions on PedsQL scores were largely comparable with the CHU9D utilities, except that gender heterogeneity was insignificant based on PedsQL ($P > .05$).

Discussion

The CHU9D has been specifically designed for application with children and adolescents to facilitate the estimation of QALYs for the economic evaluation of health care treatment and preventive programs targeted at young people. The results of this study indicate that the CHU9D-CHN demonstrates satisfactory psychometric properties for students in China.

Table 2 Distribution of PedsQL total scores across dimension levels of the CHU9D ($N=1912$)

CHU9D-CHN	<i>n</i> (%)	Mean PedsQL total score
Worried		
I don't feel worried today	1059 (55.39)	85.74
I feel a little bit worried today	467 (24.42)	78.99
I feel a bit worried today	219 (11.45)	76.80
I feel quite worried today	109 (5.70)	74.31
I feel very worried today	58 (3.03)	68.93
Sad		
I don't feel sad today	1456 (76.15)	84.15
I feel a little bit sad today	277 (14.49)	77.08
I feel a bit sad today	92 (4.81)	71.18
I feel quite sad today	45 (2.35)	72.96
I feel very sad today	42 (2.20)	69.12
Pain		
I don't have any pain today	1351 (70.66)	84.53
I have a little bit of pain today	360 (18.83)	77.94
I have a bit of pain today	126 (6.59)	72.76
I have quite a lot of pain today	56 (2.93)	69.55
I have a lot of pain today	19 (0.99)	67.47
Tired		
I don't feel tired today	1006 (52.62)	86.13
I feel a little bit tired today	592 (30.96)	79.06
I feel a bit tired today	201 (10.51)	75.31
I feel quite tired today	61 (3.19)	70.34
I feel very tired today	52 (2.72)	71.67
Annoyed		
I don't feel annoyed today	1418 (74.16)	84.33
I feel a little bit annoyed today	301 (15.74)	76.95
I feel a bit annoyed today	103 (5.39)	74.32
I feel quite annoyed today	57 (2.98)	69.87
I feel very annoyed today	33 (1.73)	67.41
Schoolwork/homework		
I have no problems with my schoolwork/homework today	1046 (54.71)	85.39
I have a few problems with my schoolwork/homework today	580 (30.33)	79.83
I have some problems with my schoolwork/homework today	213 (11.14)	74.39
I have many problems with my schoolwork/homework today	43 (2.25)	71.10
I can't do my schoolwork/homework today	30 (1.57)	70.15
Sleep		
Last night, I had no problems sleeping	1562 (81.69)	83.66
Last night, I had a few problems sleeping	205 (10.72)	76.70
Last night, I had some problems sleeping	79 (4.13)	72.29
Last night, I had many problems sleeping	29 (1.52)	66.18
Last night, I couldn't sleep at all	37 (1.94)	69.44
Daily routine		
I have no problems with my daily routine today	1697 (88.76)	83.19
I have a few problems with my daily routine today	162 (8.47)	73.72
I have some problems with my daily routine today	41 (2.14)	68.18
I have many problems with my daily routine today	6 (0.31)	66.99
I can't do my daily routine today	6 (0.31)	49.36

Table 2 (continued)

CHU9D-CHN	<i>n</i> (%)	Mean Ped-sQL total score
Able to join in activities		
I can join in with any activities today	1126 (58.89)	85.22
I can join in with most activities today	351 (18.36)	79.71
I can join in with some activities today	221 (11.56)	77.49
I can join in with a few activities today	143 (7.48)	72.90
I can join in with no activities today	71 (3.71)	72.09

Table 3 Test retest reliability

CHU9D dimensions	Percentage agreement (weighted kappa)		Test retest correlation (Spearman's Rho)		Weighted kappa coefficient	
	This study 2-week (<i>n</i> = 232)	UK study ^a , M–A (<i>n</i> = 24)	This study 2-week (<i>n</i> = 232)	UK study, M–A (<i>n</i> = 24)	This study 2-week (<i>n</i> = 232)	UK study, M–A (<i>n</i> = 24)
Worried	49.57*	76.0*	0.43*	0.49*	0.32*	0.31*
Sad	71.55*	79.2*	0.32*	0.55*	0.27*	0.32*
Pain	63.36*	84.4*	0.37*	0.49*	0.32*	0.46*
Tired	55.17*	72.9*	0.45*	0.52*	0.37*	0.34*
Annoyed	61.64*	81.3*	0.24	0.34	0.21*	0.34*
Schoolwork	65.09*	86.5*	0.64*	0.63*	0.53*	0.52*
Sleep	71.12*	84.4*	0.43*	0.70*	0.39*	0.60*
Daily routine	81.90*	86.5*	0.13*	0.57*	0.20	0.41*
Activities	65.95*	83.3*	0.41*	0.58*	0.35*	0.41*

M–A morning–afternoon

*Significant at $P < .05$

^aSource of the UK study (children aged 6–7): Canaway and Frew [41], Table 4, p. 177

The CHU9D-CHN has demonstrated an acceptable level of internal consistency and good retest reliability for evaluating the HRQoL of Chinese students aged 8–17 from Grades 3–9. Although there is one item for each dimension of the CHU9D, the nine items were tapped into the same overall construct but represent different domains [40]. The Cronbach's alpha for the CHU9D-CHN in this Chinese student sample (0.771) is similar to what has been reported in an Australian study (0.778) [40]. The retest reliability is satisfactory as measured by ICC (0.653) which is larger than 0.6. At dimension level, percentage agreements ranged from 50 (worried) to 82 (daily routine), and the Kappa coefficients ranged from poor to moderate. Comparing the results from this study with literature, it is found that the percentage agreements and kappa coefficient were generally lower than those reported with a pragmatic morning/afternoon test–retest of the CHU9D among children aged 6–7 years old

in UK [41]. The key reasons behind may be due to a much longer interval (2-week) adopted for retest in this study and the markedly lower age range in the UK study. Subjective evaluations on HRQoL could have changed significantly during 2 weeks. Unfortunately, test–retest measurements could not be carried out in the same or next day because of the actual operations of the investigation sites.

Results also support the convergent and known-groups validities of the CHU9D-CHN. The findings for the CHU9D-CHN in this Chinese student sample are similar to what has been reported in an Australian study [42]. For example, the CHU9D utility scores moderately correlated with the PedsQL total scores in both study; even though, the overall correlation coefficient in this study ($r = .57$) was slightly lower than that of the Australian study ($r = .63$); a moderate degree of correlation was found between similar dimensions in both study, e.g. 'Schoolwork' in the CHU9D

Table 4 Correlations between the CHU9D dimensions and the PedsQL dimensions

	Worried	Sad	Pain	Tired	Annoyed	Schoolwork	Sleep	Daily routine	Activities	CHU9D utility (Australian tariff)	CHU9D utility (Chinese tariff)
Physical health	0.237	0.241	0.286	0.289	0.236	0.235	0.213	0.213	0.321	0.447	0.451
Psychosocial health	0.379	0.320	0.346	0.375	0.341	0.361	0.323	0.295	0.310	0.566	0.565
Emotional function	0.372	0.318	0.329	0.370	0.346	0.292	0.361	0.290	0.297	0.554	0.556
Social function	0.261	0.264	0.221	0.208	0.249	0.254	0.232	0.248	0.265	0.411	0.412
School function	0.302	0.210	0.289	0.316	0.236	0.364	0.194	0.172	0.194	0.440	0.437
PedsQL total score	0.359	0.323	0.351	0.371	0.341	0.344	0.307	0.307	0.344	0.569	0.571

Absolute values reported. All Spearman's correlations are statistically significant (all P value $< .001$)

and 'School function' in the PedsQL. At the dimension level, the strongest degree of correlation was found between the 'Emotional function' dimension in the PedsQL and the 'Worried' dimension in CHU9D, and consistently low correlations between all the items in the CHU9D and the 'Social function' dimension in the PedsQL.

Compared to other studies conducted in developed countries, we included the investigation of the left-behind status in the known-groups validity test as this is a well-documented issue in China. As expected, the CHU9D utility scores of participants who were left-behind were significantly lower than those not left-behind, further supporting the known-groups validity (in addition to the self-reported health status and health services utilisation).

The present study has some limitations. Firstly, the survey was performed in schools, and owing to the nature of this sample, the majority of participants were healthy. As expected, based on the CHU9D-CHN descriptive system, participants generally reported a high HRQoL, with 17% reporting themselves in full health, and no reporting in the lowest or worst level of impairment for all nine dimensions of the CHU9D-CHN. This result was similar to the previous studies based on community sample [15]. Future studies should investigate the psychometric properties of the CHU9D-CHN in clinical patient samples. Secondly, future study based on patient samples should consider a longitudinal survey design to further investigate the responsiveness of the CHU9D-CHN instrument. Thirdly, the study population was limited to school children and adolescents in Shaanxi Province that participants were all of the Han ethnic group. It is unsure whether the results hold for the minorities in China. Lastly, owing to the classroom setting of questionnaire completion and the availability of researchers on site for conceptual understanding issues, it was difficult to draw a conclusion on the practicality of the instrument. It was also not possible to accurately measure the completion time for CHU9D-CHN instrument in the hard-copy questionnaire for each student. In general, students reported that the CHU9D-CHN was simple and students in Grade 5–9 typically took less than 2 min to complete the questionnaire, whilst the students in Grade 3–4 typically took 3–4 min. From an operational perspective, the CHU9D-CHN was easy to be administered, and students had little trouble answering the questions suggesting it could be implemented within a large-scale survey in the future.

Conclusion

The CHU9D-CHN is a reliable and valid instrument for measuring HRQoL of children and adolescents aged 8–17 years in China. Further studies are required to understand the psychometric properties of the CHU9D-CHN instrument in different patient samples in China.

Table 5 Comparison of the CHU9D utility score and the PedsQL total scores according to participants' characteristics

Characteristics	N (%)	CHU9D (Australian tariff)		CHU9D (Chinese tariff)		PedsQL	
		Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)
Analysing sample	1912 (100)	0.78 (0.19)	0.81 (0.67–0.92)	0.85 (0.14)	0.88 (0.78–0.96)	81.91 (12.55)	82.95 (73.91–92.39)
Gender							
Male	1066 (55.75)	0.77 (0.20)	0.80 (0.65–0.91)	0.84 (0.14)	0.87 (0.78–0.95)	82.07 (12.54)	83.69 (73.91–92.39)
Female	846 (44.25)	0.79 (0.19)	0.82 (0.69–0.96)	0.86 (0.14)	0.89 (0.79–0.96)	81.70 (12.57)	82.61 (73.91–92.04)
P value ^a		.02		.03		.52	
Age (years)							
8–12	1089 (56.96)	0.82 (0.18)	0.87 (0.73–0.99)	0.88 (0.13)	0.90 (0.81–0.10)	83.65 (12.40)	85.87 (76.08–93.48)
13–17	823 (43.04)	0.72 (0.20)	0.75 (0.61–0.88)	0.81 (0.15)	0.83 (0.74–0.91)	79.59 (12.37)	79.35 (71.74–89.13)
P value ^a		<.001		<.001		<.001	
Self-reported health							
Excellent	540 (28.24)	0.86 (0.15)	0.91 (0.79–1.00)	0.90 (0.11)	0.94 (0.85–1.00)	88.45 (10.73)	91.30 (81.52–96.74)
Very good	633 (33.11)	0.80 (0.17)	0.82 (0.70–0.92)	0.86 (0.11)	0.89 (0.80–0.96)	83.18 (10.40)	84.78 (76.08–91.30)
Good	407 (21.29)	0.73 (0.20)	0.77 (0.61–0.88)	0.82 (0.14)	0.85 (0.75–0.91)	78.56 (11.27)	78.26 (71.74–86.95)
Fair or poor	332 (17.36)	0.66 (0.22)	0.69 (0.49–0.81)	0.76 (0.16)	0.79 (0.66–0.88)	72.94 (13.86)	72.83 (64.13–83.69)
P value ^b		<.001		<.001		<.001	
Frequency of doctor visits in the last month (times)							
0	1253 (65.53)	0.80 (0.19)	0.83 (0.70–0.96)	0.86 (0.14)	0.89 (0.80–0.96)	83.14 (11.89)	83.69 (75.00–92.79)
1–2	582 (30.44)	0.75 (0.19)	0.79 (0.64–0.91)	0.83 (0.14)	0.85 (0.75–0.94)	80.23 (13.11)	81.52 (71.74–91.30)
≥ 3	77 (4.03)	0.67 (0.25)	0.73 (0.46–0.88)	0.76 (0.19)	0.80 (0.62–0.80)	74.52 (14.67)	75.00 (66.30–85.33)
P value ^a		<.001		<.001		<.001	
Left-behind children status							
Yes	1225 (64.07)	0.77 (0.20)	0.80 (0.66–0.92)	0.84 (0.14)	0.87 (0.78–0.95)	81.32 (12.89)	82.61 (73.86–92.04)
No	687 (35.93)	0.80 (0.19)	0.83 (0.69–0.96)	0.86 (0.14)	0.89 (0.79–0.96)	82.95 (11.85)	84.78 (75.00–92.39)
P value ^a		.007		.007		.015	

^aMann–Whitney *U* test^bKruskal–Wallis test

Compliance with ethical standards

Conflict of interest GC was involved in the development of the Chinese version of CHU9D instrument. All other authors declare that they have no conflict of interest.

Ethical Approval Ethical approval was granted by the Ethics Committee of Baoji Center for Disease Control and Prevention (Project No. 201601). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Written informed consent was obtained from both parents or caregivers and all individual students included in the study.

References

- Williams, P. G., Holmbeck, G. N., & Greenley, R. N. (2002). Adolescent health psychology. *Journal of Consulting and Clinical Psychology, 70*, 828–842.
- Sawyer, S. M., Afifi, R. A., Bearinger, L. H., Blakemore, S. J., Dick, B., Ezech, A. C., & Patton, G. C. (2012). Adolescence: A foundation for future health. *Lancet, 379*, 1630–1640. [https://doi.org/10.1016/s0140-6736\(12\)60072-5](https://doi.org/10.1016/s0140-6736(12)60072-5).
- Kleinert, S. (2007). Adolescent health: An opportunity not to be missed. *Lancet, 369*, 1057–1058. [https://doi.org/10.1016/s0140-6736\(07\)60374-2](https://doi.org/10.1016/s0140-6736(07)60374-2).
- Morris, J., Perez, D., & McNoe, B. (1998). The use of quality of life data in clinical practice. *Quality of Life Research, 7*, 85–91.
- Lehnert, T., Sonntag, D., Konnopka, A., Riedel-Heller, S., & König, H. H. (2012). The long-term cost-effectiveness of obesity prevention interventions: Systematic literature review. *Obesity Reviews, 13*, 537–553. <https://doi.org/10.1111/j.1467-789X.2011.00980.x>.
- Chen, G., & Ratcliffe, J. (2015). A review of the development and application of generic multi-attribute utility instruments for paediatric populations. *Pharmacoeconomics, 33*, 1013–1028. <https://doi.org/10.1007/s40273-015-0286-7>.
- Stevens, K. J. (2010). Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. *Qualitative Health Research, 20*, 340–351. <https://doi.org/10.1177/1049732309358328>.
- Stevens, K. (2009). Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Quality of Life Research, 18*, 1105–1113. <https://doi.org/10.1007/s11136-009-9524-9>.

9. Ratcliffe, J., Flynn, T., Terlich, F., Stevens, K., Brazier, J., & Sawyer, M. (2012). Developing adolescent-specific health state values for economic evaluation: An application of profile case best-worst scaling to the Child Health Utility 9D. *Pharmacoeconomics*, *30*, 713–727. <https://doi.org/10.2165/11597900-000000000-00000>.
10. Stevens, K., & Ratcliffe, J. (2012). Measuring and valuing health benefits for economic evaluation in adolescence: An assessment of the practicality and validity of the child health utility 9D in the Australian adolescent population. *Value Health*, *15*, 1092–1099. <https://doi.org/10.1016/j.jval.2012.07.011>.
11. Chen, G., Flynn, T., Stevens, K., Brazier, J., Huynh, E., Sawyer, M., Roberts, R., & Ratcliffe, J. (2015). Assessing the health-related quality of life of Australian adolescents: An empirical comparison of the Child Health Utility 9D and EQ-5D-Y instruments. *Value Health*, *18*, 432–438. <https://doi.org/10.1016/j.jval.2015.02.014>.
12. Blake, H., Quirk, H., Leighton, P., Randell, T., Greening, J., Guo, B., et al. (2016). Feasibility of an online intervention (STAK-D) to promote physical activity in children with type 1 diabetes: Protocol for a randomised controlled trial. *Trials*, *17*(1), 583. <https://doi.org/10.1186/s13063-016-1719-0>.
13. Furber, G., & Segal, L. (2015). The validity of the Child Health Utility instrument (CHU9D) as a routine outcome measure for use in child and adolescent mental health services. *Health and Quality of Life Outcomes*, *13*(1), 22. <https://doi.org/10.1186/s12955-015-0218-4>.
14. Frew, E. J., Pallan, M., Lancashire, E., Hemming, K., & Adab, P. (2015). Is utility-based quality of life associated with overweight in children? Evidence from the UK WAVES randomised controlled study. *BMC Pediatrics*, *15*(1), 211. <https://doi.org/10.1186/s12887-015-0526-1>.
15. Xu, F., Chen, G., Stevens, K., Zhou, H., Qi, S., Wang, Z., Hong, X., Chen, X., Yang, H., Wang, C., & Ratcliffe, J. (2014). Measuring and valuing health-related quality of life among children and adolescents in mainland China—A pilot study. *PLoS ONE*, *9*, e89222. <https://doi.org/10.1371/journal.pone.0089222>.
16. Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., et al. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value Health*, *8*(2), 94–104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x>.
17. Ratcliffe, J., Huynh, E., Chen, G., Stevens, K., Swait, J., Brazier, J., Sawyer, M., Roberts, R., & Flynn, T. (2016). Valuing the Child Health Utility 9D: Using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. *Social Science and Medicine*, *157*, 48–59. <https://doi.org/10.1016/j.socscimed.2016.03.042>.
18. Chen, G., Xu, F., Huynh, E., Wang, Z., Li, C., Stevens, K., & Ratcliffe, J. (2016). Scoring the Child Health Utility 9D instrument. Estimation of a Chinese adolescent-specific tariff. *Quality of Life Research*, *25*(S1), 23–24.
19. Chen, G., Huynh, E., Xu, F., Stevens, K., Brazier, J., Swait, J., & Ratcliffe, J. (2017). OP55 Health technology assessment in children and adolescents: Adolescent preferences for Child Health Utility 9D health states. *International Journal of Technology Assessment in Health Care*, *33*(S1), 24–25. <https://doi.org/10.1017/s026646231700143x>.
20. Chen, G., Xu, F., Huynh, E., Wang, Z., Li, C., Stevens, K., & Ratcliffe, J. (2018). *Developing a Chinese-specific adolescent tariff for the Child Health Utility 9D instrument*. Melbourne: Centre for Health Economics Research Paper 96, Monash University.
21. Hao, Y., Tian, Q., Lu, Y., Chai, Y., & Rao, S. (2010). Psychometric properties of the Chinese version of the Pediatric Quality of Life Inventory 4.0 generic core scales. *Quality of Life Research*, *19*, 1229–1233. <https://doi.org/10.1007/s11136-010-9672-y>.
22. Huang, Y., Zhong, X. N., Li, Q. Y., Xu, D., Zhang, X. L., Feng, C., Yang, G. X., Bo, Y. Y., & Deng, B. (2015). Health-related quality of life of the rural-China left-behind children or adolescents and influential factors: A cross-sectional study. *Health and Quality Life Outcomes*, *13*, 29. <https://doi.org/10.1186/s12955-015-0220-x>.
23. Wu, H. H., Li, H., & 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 and Quality Life Outcomes*, *11*, 87. <https://doi.org/10.1186/1477-7525-11-87>.
24. Hofsteenge, G. H., Weijs, P. J., Delemarre-van de Waal, H. A., de Wit, M., & Chinapaw, M. J. (2013). Effect of the Go4it multidisciplinary group treatment for obese adolescents on health related quality of life: A randomised controlled trial. *BMC Public Health*, *13*, 939. <https://doi.org/10.1186/1471-2458-13-939>.
25. Varni, J. W., Seid, M., & Kurtin, P. S. (2001). PedsQL™ 4.0: Reliability and validity of the Pediatric Quality of Life Inventory™ version 4.0 generic core scales in healthy and patient populations. *Medical Care*, *39*, 800–812.
26. Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*, 613–619. <https://doi.org/10.1177/001316447303300309>.
27. Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage. Publications Ltd.
28. Fayers, P. M., & Machin, D. (2007). *Quality of Life: The assessment, analysis and interpretation of patient-reported outcomes* (2nd ed). Chichester: Wiley.
29. Wang, H. M., Patrick, D. L., Edwards, T. C., Skalicky, A. M., Zeng, H. Y., & Gu, W. W. (2012). Validation of the EQ-5D in a general population sample in urban China. *Quality of Life Research*, *21*(1), 155–160. <https://doi.org/10.1007/s11136-011-9915-6>.
30. Conner-Spady, B. L., Marshall, D. A., Bohm, E., Dunbar, M. J., Loucks, L., Khudairy, A. A., et al. (2015). Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. *Quality of Life Research*, *24*(7), 1775–1784. <https://doi.org/10.1007/s11136-014-0910-6>.
31. Cheung, P. W. H., Wong, C. K. H., Samartzis, D., Luk, K. D. K., Lam, C. L. K., Cheung, K. M. C., et al. (2016). Psychometric validation of the EuroQoL 5-Dimension 5-Level (EQ-5D-5L) in Chinese patients with adolescent idiopathic scoliosis. *Scoliosis Spinal Disorders*, *11*, 19. <https://doi.org/10.1186/s13013-016-0083-x>.
32. Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213–220.
33. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.
34. Anderson, T. W., & Finn, J. (1996). *The new statistical analysis of data*. New York: Springer.
35. Fan, F., Su, L. Y., Gill, M. K., & Birmaher, B. (2010). Emotional and behavioral problems of Chinese left-behind children: A preliminary study. *Social Psychiatry Psychiatric Epidemiology*, *45*(6), 655–664. <https://doi.org/10.1007/s00127-009-0107-4>.
36. Ding, G., & Bao, Y. (2014). Editorial perspective: Assessing developmental risk in cultural context: The case of ‘left behind’ children in rural China. *Journal of Child Psychology and Psychiatry*, *55*(4), 411–412. <https://doi.org/10.1111/jcpp.12228>.
37. Man, Y., Mengmeng, L., Lezhi, L., Ting, M., & Jingping, Z. (2017). The psychological problems and related influential

- factors of left-behind adolescents (LBA) in Hunan, China: A cross sectional study. *International Journal of Equity in Health*, 16(1), 163. <https://doi.org/10.1186/s12939-017-0639-2>.
38. Qu, G. B., Wu, W., Wang, L. L., Tang, X., Sun, Y. H., Li, J., et al. (2017). Systematic review and meta-analysis found higher levels of behavioural problems in male left-behind children aged 6–11 years. *Acta Paediatrica*. <https://doi.org/10.1111/apa.14199>.
39. Xing, H., Yu, W., Xu, F., & Chen, S. (2017). Influence of social support and rearing behavior on psychosocial health in left-behind children. *Health Quality Life Outcomes*, 15(1), 13. <https://doi.org/10.1186/s12955-017-0592-1>.
40. Furber, G., & Segal, L. (2015). The validity of the Child Health Utility instrument (CHU9D) as a routine outcome measure for use in child and adolescent mental health services. *Health Quality Life Outcomes*, 13, 22. <https://doi.org/10.1186/s12955-015-0218-4>.
41. Canaway, A. G., & Frew, E. J. (2013). Measuring preference-based quality of life in children aged 6–7 years: A comparison of the performance of the CHU-9D and EQ-5D-Y—the WAVES pilot study. *Quality of Life Research*, 22(1), 173–183. <https://doi.org/10.1007/s11136-012-0119-5>.
42. Petersen, K. D., Chen, G., Mpundu-Kaambwa, C., Stevens, K., Brazier, J., & Ratcliffe, J. (2018). Measuring health-related quality of life in adolescent populations: An empirical comparison of the CHU9D and the PedsQL™ 4.0 Short Form 15. *Patient*, 11(1), 29–37. <https://doi.org/10.1007/s40271-017-0265-5>.