

RESEARCH

Open Access



# Item reduction and validation of the Chinese version of diabetes quality-of-life measure (DQOL)

Xuejing Jin<sup>1,2</sup>, Gordon G. Liu<sup>2,3</sup>, Hertz C. Gerstein<sup>4</sup>, Mitchell A. H. Levine<sup>1,8</sup>, Kathleen Steeves<sup>5</sup>, Haijing Guan<sup>2,6</sup>, Hongchao Li<sup>7</sup> and Feng Xie<sup>1\*</sup> 

## Abstract

**Background:** The Diabetes Quality-of-Life (DQOL) Measure is a 46-item diabetes-specific quality of life instrument. The original English version of the DQOL has been translated into Chinese after cultural adaptation, and the Chinese DQOL has been validated in the Chinese diabetic patient population and used in diabetes-related studies. There are two recognized problems with the Chinese DQOL: 1) the instrument is too long, and 2) the non-response rate of certain items is relatively high. This study aimed to develop and validate a short version for the Chinese DQOL.

**Methods:** Item reduction was conducted based on the classical test theory (CTT) and item response theory (IRT), each combined with exploratory factor analysis (EFA). The confirmatory factor analysis (CFA) and Spearman correlation coefficient were employed in validating the short versions.

**Results:** Both the study sample ( $n = 2,886$ ) and the validation sample ( $n = 2,286$ ) were from a longitudinal observation study of Chinese type 2 diabetic patients. The CTT kept 32 items, and the IRT kept 24 items from the original 46-item version. The two short versions were comparable in psychometric properties.

**Conclusion:** The 24-item IRT-based short version of the Chinese DQOL was selected as the preferred short version because it imposes a lower burden on patients without compromising the psychometric properties of the instrument.

**Keywords:** Item response theory, Classical test theory, Factor analysis, Diabetes, Quality of life, Psychometrics

## Background

The global prevalence of diabetes mellitus (DM) in adults was 9.1% (415 million people) in 2015, which makes DM one of the most common chronic diseases around the world [1]. Diabetes-related complications, receiving blood glucose control therapies, and dealing with hypoglycemic agents and/or insulin adverse reactions seriously affect patients' (and their family members') health-related quality of life (HRQoL) in both physical and psychological ways [2, 3]. Hence, diabetic patients' HRQoL outcomes have been increasingly recognized as valuable and essential information to obtain in the fields of clinical research and diabetes management.

Diabetic patients' HRQoL are measured by generic or diabetes-specific instruments [4]. Diabetes-specific instruments, as designed to focus on diabetes specific conditions, are more sensitive to diabetes-symptoms-related impacts on life and quality of life than generic instruments [5]. The Diabetes Quality-of-Life Measure (DQOL) is one of the most commonly used diabetes-specific instruments [6, 7]. It was developed and validated to compare two treatment regimens for chronic complications in patients with diabetes in the Diabetes Control and Complications Trial (DCCT) [8, 9]. The DQOL contains a total of 46 items, and all the items are categorized into one of the following four domains: life satisfaction (15 items), diabetes impact (20 items), social/vocational related worries (7 items), and diabetes related worries (4 items). The DQOL adopts a 5-point Likert scale for its response options. The scores range from 1, labeled as "very satisfied," to 5, labeled as "very dissatisfied," for items in the life satisfaction

\* Correspondence: [fengxie@mcmaster.ca](mailto:fengxie@mcmaster.ca)

<sup>1</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main St W, Hamilton, ON L8S 4K1, Canada  
Full list of author information is available at the end of the article



domain; from 1, labeled as “never impacted,” to 5, labeled as “always impacted,” for items in the diabetes impact domain; and from 1, labeled as “never worried,” to 5, labeled as “always worried,” for the social/vocational related and diabetes related worries domains.

The DQOL has been translated into five languages, including Chinese [10]. This measure was first translated and adapted for Chinese-Canadians who lived in the Toronto area by Cheng et al. [11, 12]. They removed 10 privacy-related (e.g. sexual life) items from the original DQOL and added six items regarding diet, worrying about death and so on. However, there was not sufficient psychometric evidence to support the cultural adaptation in Cheng et al.’s study [11], and the translation and validation were conducted based on an immigrant population, which cannot necessarily be generalized to the entire Chinese diabetic patient population. Ding et al. translated and adapted the DQOL for the Chinese population based on a sample of Chinese patients with diabetes who lived in Mainland China [13], and conducted validation of the Chinese DQOL on a separate sample of Chinese patients with type 2 DM living in Mainland China [14]. The wording of seven items was changed in Ding et al.’s adaptation (see Additional file 1). Currently, the Chinese DQOL translated and adapted by Ding et al. has been used in diabetes-related clinical studies in China [15–17]. During its application among the Chinese diabetic patient population, the Chinese DQOL has exposed some of its own issues [18]. First, the non-response rate of certain privacy-related items was relatively high; and second, interviewees complained that the instrument was too long [19, 20]. In order to solve these issues, developing and validating a short version of the Chinese DQOL is necessary.

The classical test theory (CTT) and the item response theory (IRT) are two commonly used psychometric theories in conducting item selection and reduction for measures; however, these two theories work based on different assumptions and statistical approaches, and both have shortcomings [21, 22]. More specifically, the CTT assumes that each respondent has a true total score,  $T$  (latent variable), and each item is a representative of the score  $T$ ; while the IRT follows the assumptions that the latent trait of a measure is unidimensional and all items are conditionally independent of each other. Generally, CTT tests the difficulty and discrimination at the item level and the reliability at the whole measure level; while IRT uses a set of logistic regression models to estimate the “discrimination,” “location,” and “information” for each item [21, 22]. The CTT is limited by the sample and item/test dependence and equal error (of measurement across examinees) assumption [21, 22]. The IRT overcomes these shortcomings but requires for large sample sizes for model fitting [21, 22]. There is no

generally accepted approach or standard for item reduction. Currently, researchers have been using the IRT alone [23], the combination of the IRT and factor analyses [24, 25], or the combination of the CTT and factor analysis [26, 27] when selecting or reducing items.

Therefore, the present study aims to use both the CTT and IRT combined with factor analysis to derive and validate a short version of the Chinese DQOL, which can be rapidly administered in practice and can reduce response burden on patients.

## Methods

### Sample and data

We used the data from a Chinese community-based longitudinal survey of clinically diagnosed type 2 diabetic patients (T2DP) from five cities: Beijing, Chengdu, Guangzhou, Nanjing, and Shenyang. Patients were recruited and interviewed between December 2010 and October 2011, and followed every three months over a one-year study period. The Chinese DQOL and the EQ-5D-3L were administered at the baseline and at 12-months. Demographic, social-economic and diabetic-related information was also collected. We used the baseline data as the study sample for item reduction analysis, and the one-year end follow-up data as the validation sample to test the short versions of the Chinese DQOL reduced by CTT and IRT.

### Reduction based on the classical test theory

Three steps were used to reduce the number of items based on the CTT. The first step tested each item at the individual item level, and the second and third steps examined the items at the whole measure or domain level. The following provides the details of the tests in each step and the corresponding item removal criteria.

#### Step 1. Item level tests

We tested three item level properties for each of the 46 items in this step, i.e., missing rate, item score mean, and item score standard deviation (SD).

Items which are unclear, ambiguous, or potentially embarrassing usually have a higher chance to have high non-response rate issues. This kind of item can provide very limited useful information, and its results are hard to interpret [21]. The exclusion criterion for the missing rate was higher than 5% [28].

In the CTT, item difficulty and discrimination are often evaluated in item level testing; however, most of the item difficulty and discrimination indexes are designed to test dichotomous items and can hardly be applied to test Likert items [29]. Norman has provided compelling evidence on the appropriateness of using descriptive statistics and parametric methods to test Likert items [30, 31]. The mean and SD of an item can provide

fundamental information on whether the item can provide useful information or not [32]. For example, if the mean score is 4.7 for a 5-point Likert item (score range: 1 to 5), then the item is left-skewed and may not be able to provide the information it was designed to collect. In addition, if the SD of an item is low, then the item has low variability and it may not be useful either. There are no generally accepted criteria for the item level test using mean and standard deviation, and we used the most lenient criteria reported in the existing studies. We used the lowest score option plus 20% of the score range and the highest score option minus 20% of the score range to define the cut point of the exclusion criterion in terms of item score mean [21, 33, 34]. The lowest and highest score options for each item is 1 and 5, respectively, and the score range for each item is 4. Thus, the exclusion criterion for the item score mean was lower than 1.8 or higher than 4.2. The exclusion criterion for the item score SD was smaller than one-sixth of the score range, i.e., 0.67 ( $1/6 \times 4$ ) [21, 33–35].

Any item that met any two or more of the three exclusion criteria was removed from the measure. In addition, any item with a missing rate higher than 10% was removed regardless of the results of the other two criteria.

### **Step 2. Exploratory factor analysis**

In this step, exploratory factor analysis (EFA) was employed on the remaining items to examine the underlying structure of the measure and remove items with low factor loadings on common factors.

More specifically, Bartlett's test of sphericity [36] and Kaiser-Meyer-Olkin's (KMO) measure of sampling adequacy [37] were conducted before conducting the EFA. Since the training sample violated the assumption of multivariate normality, we employed the principal-factor extraction method [38]. A scree plot was used to identify the number of factors [39]. Oblique rotation method was used in the EFA since the DQOL items were not completely unrelated to each other [40]. In this step, any item with a factor loading less than 0.3 was removed [41].

### **Step 3. Internal consistency reliability**

Internal consistency reliability was tested in terms of the corrected item-total correlation and Cronbach's alpha [29]. Both tests were conducted at the factor level based on the results of the EFA in step 2.

Since there is no standard scoring method for the Chinese DQOL, we used the patients' mean score of the items in each factor as the "factor score" when calculating the corrected item-total correlation. For each item, the corrected item-total correlation was calculated as the Pearson correlation coefficient between the item score and the mean score of the rest of the items in the factor this item belonged to. A larger corrected item-total correlation

coefficient indicates better internal consistency reliability. The exclusion criterion was a correlation coefficient smaller than 0.3 [42]. For the Cronbach's alpha, the exclusion criterion was that the Cronbach's alpha of the factor increased after removing an item [43].

In this step, any item that met one or more of these two exclusion criteria was removed from the measure. An additional EFA was used to check if the factor structure changed after this step; if so, the new factor structure would be used as the final structure of the short version developed based on the CTT.

### **Reduction based on the item response theory**

One of the basic assumptions of the IRT is unidimensionality [44]; however, DQOL was designed to measure multiple aspects of burden that diabetes places on patients. In order to conduct the IRT analysis without violating the assumption of unidimensionality, we employed EFA in the first place to re-identify the potential dimensional structure of the original Chinese DQOL and then fitted the sets of IRT models for each individual dimension. Details of the two steps are as follows.

#### **Step 1. Exploratory factor analysis**

Similar to the EFA analysis process adopted under the CTT reduction approach, Bartlett's test and KMO test were carried out for testing the sphericity and sampling adequacy, respectively, before implementing the EFA under the IRT reduction approach. Number of factors was identified by a scree plot generated based on the 46 Chinese DQOL items. Then principal-factor extraction method and oblique rotation method were employed to conduct the EFA. In this step, any item with a factor loading of less than 0.3 was removed.

#### **Step 2. Item response theory analysis**

The graded response model (GRM), which is a type of item response model for items with ordered response options [45], was employed in this step to analyze the remaining items within each factor identified in step 1. The GRM was first introduced by Samejima [45]. It models each item with its own discrimination parameter and a set of parameters that identify the boundaries between the ordered options using a logistic regression approach. The item information functions (IIFs) were built based on the fitted GRMs to evaluate the "information", i.e., reliability, each item contributed to the factor. A larger amount of information an item can provide indicates a better item it is. The GRM and IIF formulas are presented in the [Appendix](#).

In this step, any item that had an estimation of discrimination parameter less than 1.0 [46] and provided item information less than 0.5 was removed from the measure [25]. An additional EFA was also conducted to

check the factor structure; and if the structure changed after this step, the new factor structure would be used as the final structure of the short version developed based on the IRT.

### **Validating and comparing the two short versions of the Chinese DQOL**

We evaluated and compared the two short versions at three aspects, i.e., performance in the confirmatory analysis (CFA), correlation with EQ-5D, and the magnitude of reduced response burden.

#### **Confirmatory factor analysis**

The CFA was employed to validate the structure of the two short versions of the Chinese DQOL. We specified that the domains were correlated with each other and employed maximum likelihood estimation in the CFA. Two statistics produced by the CFA were used to compare the performance of the two versions: standardized root mean squared residual (SRMR) and comparative fit index (CFI).

The SRMR is the square root of the difference between the residuals of the sample covariance matrix and the proposed covariance model. It ranges from 0 to 1, and a smaller value indicates a better fit [47]. The CFI compares the sample covariance matrix with this null model based on the assumption that all latent variables (factors) are uncorrelated. The CFI ranges from 0 to 1, and a larger value indicates a better fit [47]. Since the variation of performance among fit indices, according to Hu and Bentler's two-index presentation strategy [48], we adopted the SRMR as the fundamental fit index and the CFI and as a supplementary index.

#### **Correlation with the EQ-5D**

We tested the construct validity of the two reduced versions of the Chinese DQOL against the EQ-5D-3L index and EQ visual analogue scale (EQ-VAS).

The EQ-5D-3L is a widely used preference-based generic quality of life instrument which has 5 questions that ask about whether there are any problems in: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each question has three response levels, i.e., no problems, some (or moderate) problems, and extreme problems (or unable to). Patients' EQ-5D-3L responses were converted in to EQ-5D-3L values by using the Chinese EQ-5D-3L value set [49]. The EQ-VAS records the patient's self-rated health on a vertical, visual analogue scale which ranges from 0 (the worst imaginable health state) to 100 (the best imaginable health state) [50].

Spearman's correlation coefficients between the EQ-5D-3L index and the mean score of each one of the two short versions of the Chinese DQOL were calculated respectively. The correlation coefficients between the EQ-VAS and the two short versions were also calculated

individually. A larger correlation coefficient indicates a higher construct validity [28, 29].

#### **Final short version selection**

The short version which performed better in both the CFA and had higher correlation with EQ-5D was selected as the final short version of the Chinese DQOL. In the event of any conflict between the CFA and the correlation analysis results, we selected the short version reduced more response burden as the final short version of the Chinese DQOL.

All statistical analyses were conducted with a two-tailed test at the significance level of 0.05 in STATA 14.2 (StataCorp LP, Texas, USA).

## **Results**

### **Sample**

A total of 2886 patients were recruited and interviewed at the baseline. The mean age and diabetes duration of the study sample was 61.15 years and 7.94 years, respectively. Among all patients, 55.68% were female, 64.10% were retired, and 16.18% had used insulin in the last 6 months. The mean scores of the EQ-5D-3L index, VAS, and the Chinese DQOL (mean score of the 46 items) were 0.89, 72.71, and 2.07, respectively (Table 1). In the validation analyses, the CFA and the calculation of the EQ-5D-3L index only employed observations without missing data. Because of this, our validation sample only included patients with no missing values on responses to the 5 questions of the EQ-5D and to the DQOL items kept after the item reduction based on the CTT and IRT. Of the 2542 patients who completed the year-end follow-up, 2286 were included in the validation sample (Table 1). Compared to the study sample, the validation sample had a higher proportion of people who were older, retired, and used insulin (Table 1).

### **Item reduction results**

Tables 2 and 3 show the item reduction results based on the CTT and IRT, respectively. A total of 14 and a total of 22 items (details see supplementary materials) were removed from the Chinese DQOL based on the CTT and IRT, respectively.

In step 1 of the reduction based on the CTT, two items, item #10 (satisfied with sex life) and item #25 (interferes with sex life) were removed from the measure because their missing rates were higher than 10%. Item #32 (being teased because of having diabetes), item #36 (worry about marriage), item #40 (worry about completing education), and item #41 (worry about unemployment) were removed because of their low mean scores (all < 1.8) and small SDs (all < 0.67). Item #35 (hide having an insulin reaction) was removed because of the high missing rate (8.07%) and low mean score and small SD.

**Table 1** Patients' baseline demographic and diabetes-related information

	Study sample (N = 2886)	Validation sample (N = 2286 <sup>a</sup> )	P-value <sup>b</sup>
Age mean (SD), years	61.15 (11.42)	61.84 (11.25)	0.030
Female	1607 (55.68)	1269 (55.51)	0.910
Diabetes duration mean (SD), years	7.94 (6.75)	7.69 (6.33)	0.174
Retired	1850 (64.10)	1544 (67.54)	0.010
Used insulin in the last 6 months	467 (16.18)	476 (20.82)	< 0.001
Controlled diet	2580 (89.40)	2036 (89.06)	0.718
EQ-5D-3 L index mean (SD)	0.89 (0.14)	0.89 (0.13)	0.999
EQ-ED VAS mean (SD)	72.71 (15.46)	73.00 (15.48)	0.503
DQOL score mean (SD) <sup>c</sup>	2.07 (0.38)	2.07 (0.39)	0.999

The values presented are numbers (percentage) unless otherwise stated

<sup>a</sup>A total of 2542 patients finished the last round follow-up at the year end. Only observations with no missing data on the EQ-5D-3L and DQOL questions were included in the validation sample

<sup>b</sup>T-test for mean and Chi-square test for frequency

<sup>c</sup>The DQOL score was calculated from the mean of the 46 items

SD Standard deviation; VAS Visual analogue scale; DQOL Diabetes quality-of-life measure

In step 2, the EFA identified two factors among the remaining items. Item #7 (satisfied with knowledge about diabetes), item #23 (feel good about yourself), item #26 (interfere with riding a bike or using a machine), item #29 (explain what it means to have diabetes), item #31 (tell others about your diabetes), and item #34 (eat something you shouldn't rather than tell someone that you have diabetes) were removed due to low factor loadings (< 0.3). In step 3, item #38 (worry about whether you can get a job you want) was removed because of the low correlation with the mean score of the factor it belonged to. The factor structure identified in Step 2 remained the same after removing item #38 in Step 3.

In the reduction based on the IRT, the EFA identified 2 factors of the 46 DQOL items, and removed items #7, #23, #26, #29, #31, and #34 because their factor loading were all smaller than 0.3. In step 2, item #5 (satisfied with the flexibility of the diet), item #8 (satisfied with sleep), item #10, item #12 (satisfied with the appearance of your body), item #13 (satisfied with the time spent on exercising), item #18 (low blood sugar reactions), item #21 (bad night's sleep), item #24 (feel restricted by diet), item #25, item #32, item #33 (feel that because of diabetes you go to the bathroom more than others), item #38, item #39 (worry about the pension), item #40, and item #41 were removed in the IRT analysis due to their item discrimination being smaller than 1 and their item information being lower than 0.5 (Table 3). The factor structure identified in the EFA remained the same after the IRT analysis.

### Validation results

Table 4 shows the validation results of the two short versions of the Chinese DQOL. In the CFA, the two short versions had similar SRMRs (0.078, after rounding, for

both short versions) which were also similar to that of the original Chinese DQOL (SRMR = 0.077). The short version based on the IRT had a larger CFI (0.726) than that of the version reduced based on the CTT (CFI = 0.630). The CFI of each short versions was larger than that of the original Chinese DQOL (CFI = 0.616).

The absolute Spearman's correlation coefficient between the CTT reduced version of the DQOL and the EQ-5D-3L index scores was 0.298, which was higher than that ( $\rho = 0.288$ ) between the IRT reduced version and the EQ-5D-3L index scores. Both reduced versions had a higher correlation with the EQ-5D-3L index scores than the original Chinese DQOL ( $\rho = 0.276$ ). In terms of testing using the EQ-VAS, the CTT-based short version had a higher correlation ( $\rho = 0.288$ ) than the original version ( $\rho = 0.273$ ), and the IRT-based short version had a slightly lower correlation ( $\rho = 0.269$ ) than the original version.

### Discussion

This study shortened the 46-item Chinese version of the DQOL based on two psychometric theories, the CTT and IRT, each combined with the EFA, respectively. The two short versions were validated using the CFA and Spearman correlation coefficients. The CTT provided a short version of the Chinese DQOL with 32 items kept, and the IRT provided a short version with 24 items kept. Among the 14 items removed based on the CTT, 13 were removed based on the IRT as well.

There are few published studies we can compare our results with. Two items related to sexual life had high missing rates in our study, and were removed from the measure in the reduction processes based on both the CTT and IRT. This was consistent with the translation and cultural adaptation study conducted in 1999 among Chinese diabetic patients lived in Canada [12]. The high

**Table 2** Item reduction results based on the CTT

Item No.	Step 1. Item level tests			Step 2. EFA		Step 3. Internal consistency reliability	
	Missing rate (%)	Mean score	SD	Factor 1	Factor 2	Item-total corrected correlation coefficient <sup>b</sup>	Cronbach's alpha <sup>b</sup>
	Missing rate > 5% <sup>a</sup>	< 1.8 or > 4.2 <sup>a</sup>	< 0.67 <sup>a</sup>	Factor loading < 0.3 <sup>a</sup>		$r < 0.3^a$	Increased after remove <sup>a</sup> (Factor 1: 0.884 Factor 2: 0.822)
1	0.312	2.081	0.711	0.384	0.528 <sup>c</sup>	0.667	0.764
2	0.277	2.132	0.693	0.337	0.519 <sup>c</sup>	0.648	0.772
3	0.243	2.078	<b>0.649</b>	0.357	0.550 <sup>c</sup>	0.695	0.751
4	0.104	2.174	0.725	0.388	0.448 <sup>c</sup>	0.564	0.811
5	0.104	2.653	0.926	0.394 <sup>c</sup>	0.172	0.348	0.882
6	0.069	3.129	0.994	0.518 <sup>c</sup>	0.015	0.492	0.879
7	0.069	2.563	0.811	<b>0.297</b>	<b>0.170</b>	Removed	
8	0.035	2.740	1.090	0.398 <sup>c</sup>	0.133	0.367	0.882
9	0.139	2.112	0.679	0.420 <sup>c</sup>	0.262	0.360	0.882
10	<b>14.969</b>	2.449	0.769	Removed			
11	0.485	2.161	0.692	0.485 <sup>c</sup>	0.234	0.419	0.880
12	0.035	2.594	0.912	0.406 <sup>c</sup>	0.106	0.361	0.882
13	0.312	2.419	0.900	0.398 <sup>c</sup>	0.171	0.345	0.882
14	0.277	2.290	0.748	0.492 <sup>c</sup>	0.236	0.434	0.880
15	0.035	2.158	<b>0.656</b>	0.566 <sup>c</sup>	0.266	0.498	0.879
16	0.139	2.466	1.201	0.557 <sup>c</sup>	-0.137	0.539	0.878
17	0.069	<b>1.727</b>	0.953	0.431 <sup>c</sup>	-0.166	0.415	0.881
18	0.035	2.321	1.072	0.381 <sup>c</sup>	-0.068	0.379	0.882
19	0.035	2.556	1.075	0.534 <sup>c</sup>	-0.018	0.520	0.879
20	0.069	2.580	1.223	0.608 <sup>c</sup>	-0.165	0.597	0.877
21	0.035	2.631	1.225	0.423 <sup>c</sup>	0.060	0.396	0.882
22	0.069	<b>1.724</b>	0.892	0.492 <sup>c</sup>	-0.124	0.468	0.880
23	0.035	2.729	1.135	<b>0.230</b>	<b>0.006</b>	Removed	
24	0.035	2.960	1.239	0.381 <sup>c</sup>	-0.132	0.381	0.882
25	<b>14.414</b>	1.875	0.968	Removed			
26	1.802	<b>1.581</b>	0.961	<b>0.245</b>	<b>-0.173</b>	Removed	
27	0.416	<b>1.790</b>	1.012	0.487 <sup>c</sup>	-0.144	0.451	0.880
28	0.104	1.843	1.047	0.514 <sup>c</sup>	-0.152	0.490	0.879
29	0.035	2.445	1.176	<b>0.185</b>	<b>-0.228</b>	Removed	
30	0.208	1.882	0.977	0.574 <sup>c</sup>	-0.192	0.526	0.878
31	0.035	2.343	1.079	<b>0.222</b>	<b>-0.187</b>	Removed	
32	0.069	<b>1.297</b>	<b>0.594</b>	Removed			
33	0.104	2.499	1.260	0.355 <sup>c</sup>	-0.089	0.342	0.883
34	0.035	<b>1.781</b>	0.945	<b>0.234</b>	<b>-0.029</b>	Removed	
35	<b>8.073</b>	<b>1.136</b>	<b>0.415</b>	Removed			
36	2.495	<b>1.229</b>	<b>0.593</b>	Removed			
37	0.728	1.888	1.147	0.429 <sup>c</sup>	-0.166	0.420	0.881
38	3.222	<b>1.262</b>	0.685	0.314 <sup>c</sup>	-0.177	<b>0.294</b>	0.883
39	2.668	<b>1.311</b>	0.762	0.332 <sup>c</sup>	-0.197	0.305	0.883
40	4.089	<b>1.134</b>	<b>0.431</b>	Removed			
41	3.915	<b>1.213</b>	<b>0.633</b>	Removed			
42	0.312	<b>1.641</b>	1.018	0.424 <sup>c</sup>	-0.263	0.414	0.881

**Table 2** Item reduction results based on the CTT (Continued)

Item No.	Step 1. Item level tests			Step 2. EFA		Step 3. Internal consistency reliability	
	Missing rate (%)	Mean score	SD	Factor 1	Factor 2	Item-total corrected correlation coefficient <sup>b</sup>	Cronbach's alpha <sup>b</sup>
	Missing rate > 5% <sup>a</sup>	< 1.8 or > 4.2 <sup>a</sup>	< 0.67 <sup>a</sup>	Factor loading < 0.3 <sup>a</sup>		$r < 0.3^a$	Increased after remove <sup>a</sup> (Factor 1: 0.884 Factor 2: 0.822)
43	0.104	1.998	1.151	0.519 <sup>c</sup>	-0.264	0.508	0.879
44	0.069	<b>1.551</b>	0.908	0.441 <sup>c</sup>	-0.225	0.425	0.881
45	0.104	2.632	1.317	0.502 <sup>c</sup>	-0.182	0.487	0.880
46	0.104	<b>1.353</b>	0.720	0.427 <sup>c</sup>	-0.250	0.414	0.881

Bold and italic number indicates the item failed the corresponding test

Dashed box indicates the item(s) was removed from the scale

<sup>a</sup>Exclusion criteria

<sup>b</sup>Total scores were calculated as the corrected mean score of the factor

<sup>c</sup>indicates which factor the item belongs to based on the EFA

CTT Classical test theory, EFA Exploratory factor analysis, SD Standard deviation, N/A Not applicable

missing rate of the sexual life items is still in line with the findings in translation and cultural adaptation studies published after 2015 in other disease specific measures among the Chinese population [51]. Chinese people, especially those who are middle-aged and elderly, tend to be hesitant to talk about sex-related topics because of their relatively conservative culture background [52].

Three working and education-related items, i.e., items #38, #40, and #41, had low mean scores (Table 2) and low discriminations (Table 3), and were removed based on both the CTT and IRT. This was because most patients (64.10%) in our training sample were retired, and were not worried about working and education-related issues. These items were also removed according to the expert advice in Cheng's [11, 12] translation and cultural adaptation study.

The insulin reaction item (item #35) was removed based on both the CTT and IRT. This was because the majority of the patients in the study sample had not used insulin in the last 6 months. Similarly, the diet-related item (item #34) was also removed mainly because the majority of the patients in the study sample controlled their diet by eating healthy food and balancing their amount of food intake due to their diabetes.

In Ding's [13] translation and cultural adaptation analysis, the descriptive of item 26, "How often does your diabetes keep you from driving a car or using a machine (e.g., a typewriter)?" was changed into "How often does your diabetes keep you from riding a bike or being a typist?" This item was removed because of low factor loading in both reduction processes. Ding et al. changed the "driving a car" into "riding a bike" because civilian vehicle ownership in China was relatively low in the 1990's, and bicycles were the main means of transportation for ordinary people. However, civilian vehicle ownership in 2012 increased by 544% from 1999 [53], which may make this change in descriptive out-of-date. In addition, typewriters have long been replaced by laptops

and other smart electronics which are indispensable in contemporary Chinese people's daily lives. Therefore, further studies examining the performance of a more up-to-date descriptive, for example, "How often does your diabetes keep you from driving a vehicle or using a computer or smart phone?" are necessary.

There were 9 items that were removed in the IRT-based short version but kept in the CTT-based short version. All of these items were removed due to their low estimated discrimination and item information in the IRT analysis. One of the possible reasons for this difference is that the reduction results were impacted by the exclusion criteria we employed. Even though we used the most lenient fail criteria reported in existing studies for each, respectively, the item reduction results may still not be comparable due to the different statistical approaches applied in the two different theories.

Items #1 to #4 (satisfaction level of "the amount of time it takes to manage your diabetes," "the amount of time you spend getting a checkup," "the time it takes to determine your sugar level," and "your current treatment") were the only four treatment and diabetes management related items in the DQOL. These items loaded onto the same factor in our EFA. The rest of the 28 items in the CTT-based short version and the rest of the 20 items in the IRT-based short version belonged to the other factor, respectively. This was different than the original Chinese DQOL which has four domains. The CFA and correlation coefficients showed that the structures of the two short versions were comparable to the original version. In addition, we did not emphasize the name of the factors identified in the short versions since the present study aimed to focus on reducing the number of items for the Chinese DQOL. Content and face validity of the short versions should be examined in further studies to optimize the structure and rename the factors of the short versions.

The often-used fit indexes in the CFA are the Chi-square test and the root mean square error of approximation

**Table 3** Item reduction results based on the IRT

Item No.	Step 1. EFA		Step 2. IRT	
	Factor 1	Factor 2	Discrimination	Item information function <sup>b</sup>
	Factor loading < 0.3 <sup>a</sup>		< 1 <sup>a</sup>	< 0.5 <sup>a</sup>
1	0.356	0.421 <sup>c</sup>	2.830	> 2, < 2.5
2	0.313	0.405 <sup>c</sup>	2.848	> 2, < 2.5
3	0.342	0.407 <sup>c</sup>	3.496	> 3, < 3.5
4	0.357	0.368 <sup>c</sup>	1.987	> 1, < 1.5
5	0.361 <sup>c</sup>	0.248	<b>0.778</b>	<b>&lt; 0.2</b>
6	0.506 <sup>c</sup>	0.115	1.097	< 0.4
7	<b>0.274</b>	<b>0.181</b>	Removed	
8	0.369 <sup>c</sup>	0.258	<b>0.769</b>	<b>&lt; 0.2</b>
9	0.427 <sup>c</sup>	0.203	1.024	< 0.3
10	0.343 <sup>c</sup>	0.153	<b>0.768</b>	<b>&lt; 0.2</b>
11	0.470 <sup>c</sup>	0.160	1.211	< 0.5
12	0.365 <sup>c</sup>	0.184	<b>0.854</b>	<b>&lt; 0.3</b>
13	0.373 <sup>c</sup>	0.171	<b>0.884</b>	<b>&lt; 0.3</b>
14	0.466 <sup>c</sup>	0.218	1.147	< 0.4
15	0.544 <sup>c</sup>	0.207	1.485	> 0.6, < 0.7
16	0.544 <sup>c</sup>	0.045	1.328	> 0.5, < 0.6
17	0.456 <sup>c</sup>	-0.128	1.237	< 0.5
18	0.368 <sup>c</sup>	0.077	<b>0.801</b>	<b>&lt; 0.2</b>
19	0.494 <sup>c</sup>	0.186	1.178	< 0.5
20	0.592 <sup>c</sup>	0.031	1.567	< 0.5
21	0.394 <sup>c</sup>	0.205	<b>0.850</b>	<b>&lt; 0.3</b>
22	0.491 <sup>c</sup>	-0.063	1.506	> 0.6, < 0.7
23	<b>0.220</b>	<b>-0.023</b>	Removed	
24	0.352 <sup>c</sup>	0.068	<b>0.793</b>	<b>&lt; 0.2</b>
25	0.404 <sup>c</sup>	-0.076	<b>0.991</b>	<b>&lt; 0.3</b>
26	<b>0.262</b>	<b>-0.150</b>	Removed	
27	0.485 <sup>c</sup>	-0.052	1.429	> 0.6, < 0.7
28	0.519 <sup>c</sup>	-0.041	1.517	> 0.7, < 0.8
29	<b>0.173</b>	<b>-0.059</b>	Removed	
30	0.575 <sup>c</sup>	-0.087	1.666	> 0.8, < 0.9
31	<b>0.212</b>	<b>-0.017</b>	Removed	
32	0.346 <sup>c</sup>	-0.242	<b>0.998</b>	<b>&lt; 0.3</b>
33	0.345 <sup>c</sup>	0.042	<b>0.762</b>	<b>&lt; 0.2</b>
34	<b>0.271</b>	<b>-0.060</b>	Removed	
35	<b>0.214</b>	<b>-0.197</b>	Removed	
36	0.383 <sup>c</sup>	-0.351	1.068	< 0.4
37	0.449 <sup>c</sup>	-0.162	1.099	< 0.4
38	0.410	-0.514 <sup>c</sup>	<b>0.321</b>	<b>around 0</b>
39	0.403	-0.441 <sup>c</sup>	<b>0.303</b>	<b>around 0</b>
40	0.330	-0.461 <sup>c</sup>	<b>0.360</b>	<b>around 0</b>
41	0.372	-0.510 <sup>c</sup>	<b>0.289</b>	<b>around 0</b>
42	0.433 <sup>c</sup>	-0.182	1.224	< 0.5

**Table 3** Item reduction results based on the IRT (Continued)

Item No.	Step 1. EFA		Step 2. IRT	
	Factor 1	Factor 2	Discrimination	Item information function <sup>b</sup>
	Factor loading < 0.3 <sup>a</sup>		< 1 <sup>a</sup>	< 0.5 <sup>a</sup>
43	0.526 <sup>c</sup>	-0.089	1.357	> 0.5, < 0.6
44	0.450 <sup>c</sup>	-0.180	1.306	> 0.5, < 0.6
45	0.479 <sup>c</sup>	0.021	1.048	< 0.4
46	0.465 <sup>c</sup>	-0.280	1.515	> 0.7, < 0.8

Bold and italic number indicates the item failed the corresponding test  
Dashed box indicates the item(s) was removed from the scale

<sup>a</sup>Exclusion criteria;

<sup>b</sup>Highest point on the item information function curve;

<sup>c</sup>indicates which factor the item belongs to base on the EFA

IRT Item response theory, EFA Exploratory factor analysis

(RMSEA) [47]. In the present study, we employed the SRMR and CFI instead of the Chi-square test and RMSEA. The Chi-square test result is affected by the number of parameters, complexity of the model, and the sample size [54]. Adding more parameters into the model can improve the RMSEA as well [55]. Our two short versions of the Chinese DQOL had different numbers of items; therefore, the Chi-square test and RMSEA were inappropriate to use for comparing the CFA results of these two short versions. The SRMR is not affected by the model complexity and the number of parameters. The CFI is affected by the number of parameters added into a model, but is relatively more stable than the Chi-square test and RMSEA.

Because the two short versions of the Chinese DQOL were comparable in the validation analysis, and we did not have a hierarchy in these two criteria, we selected the short version based on the IRT (24 items) as a preferred short version for two other reasons. First, this shorter version imposes a lower burden on patients without compromising its measurement properties [56]. Second, theoretically, as a modeling statistic approach, the parameters estimated from a set of IRFs can be generalized to the entire population the study sample comes from; however, as a person statistic approach, all CTT test results can only be specified to the given study sample [57].

There are some limitations in our study. First, the training and validation samples were not independent. We did not have a truly external validation sample for our study. Second, our training sample only contained community-based patients, and most of them did not use insulin. This sample was relatively healthier than the diabetic population who had more comorbidities, was inpatient, or using insulin; therefore, our results cannot necessarily be generalized to the entire diabetic patient population. At the validation stage of this study, the CFI value of both versions did not meet the generally accepted criteria for good fit, i.e., CFI > 0.90 [47]. Even though the CFI was used as a supplementary index to



**Table 4** Validation result<sup>a</sup>

	Confirmatory factor analysis		Spearman correlation coefficient	
	SRMR	CFI	ρ(EQ-5D-3L)	ρ(EQ-VAS)
Short version based on the CTT (32 items)	0.078	0.630	-0.298	-0.288
Domain 1			-0.260	-0.260
Domain 2			-0.069	-0.148
Short version based on the IRT (24 items)	0.078	0.726	-0.288	-0.269
Domain 1			-0.240	-0.242
Domain 2			-0.069	-0.148
Original Chinese DQOL (46 items)	0.077	0.616	-0.276	-0.273

<sup>a</sup>Calculations based on a total of 1350 observations without missing values on all the five EQ-5D-3 L questions and all the 46 DQOL items  
 CTT Classical test theory, IRT Item response theory, DQOL Diabetes quality-of-life measure, SRMR Standardized root mean squared residual, CFI Comparative fit index, ρ Spearman's correlation coefficient, VAS Visual analogue scale

evaluate the model fit, this result still added uncertainty to our conclusions. Other psychometric properties such as test-retest reliability of the short version of the Chinese DQOL need to be examined in future studies.

**Conclusions**

The version developed based on the IRT retained 24 items was selected as our preferred short version of the 46-item Chinese DQOL. It can impose a lower response burden on patients in practice without compromising the psychometric properties. Further research validating the IRT-based short version of Chinese DQOL is needed.

**Appendix**

**The GRM and IIF models**

The probability of respondent *j* with latent ability level  $\theta_j$  (the latent trait for respondent *j*) to choose response option *k* or higher (in our case,  $k = 0, 1, 2, 3, 4, 5$ ) for item *i* is [45, 58]:

$$Pr(Y_{ij} \geq k | \theta_j) = \frac{\exp\{\alpha_i(\theta_j - b_{ik})\}}{1 + \exp\{\alpha_i(\theta_j - b_{ik})\}} \theta_j \sim N(0, 1)$$

where,  $\alpha_i$  represents the discrimination of item *i*, and  $b_{ik}$  is the cut-point of boundaries between the  $k^{th}$  and  $(k + 1)^{th}$  options for item *i*, which can be considered as the difficulty of choosing option *k* or higher for item *i* [45, 58].

The information function  $I_i(\theta)$  for item *i* is:

$$I_i(\theta) = \sum_{k=1}^K I_{ik}(\theta) p_{ik}(\theta)$$

where,  $I_{ik}(\theta)$  is the information function, for response option *k* of item *i*, which is defined as:

$$I_{ik}(\theta) = - \frac{\partial^2 \log p_{ik}(\theta)}{\partial \theta^2}$$

where,  $\partial$  is the partial derivative symbol, and  $p_{ik}(\theta)$  is the probability of a respondent with the latent trait level  $\theta$

choosing response option *k*, which depends on the GRM for item *i*.

**Additional file**

**Additional file 1:** Original Chinese DQOL and short versions based on the CTT and IRT. (DOCX 29 kb)

**Abbreviations**

CFA: Confirmatory factor analysis; CFI: Comparative fit index; CTT: Classical test theory; DCCT: Diabetes control and complications trial; DM: Diabetes mellitus; DQOL: Diabetes quality-of-life measure; EFA: Exploratory factor analysis; GRM: Graded response model; HRQoL: Health-related quality of life; IIF: Item information function; IRF: Item response function; IRT: Item response theory; RMSEA: Root mean square error of approximation; SD: Standard deviation; SRMR: Standardized root mean squared residual; T2DP: Type 2 diabetic patients; VAS: Visual analogue scale

**Funding**

This study used data from a survey sponsored by Guangzhou Zhongyi Pharmaceutical Co Ltd.

**Availability of data and materials**

The data that support the findings of this study are available from Gordon Liu upon reasonable request.

**Authors' contributions**

GL was the PI of the survey. GL and HL designed the protocol of the survey. XJ, HG, and HL were interviewers during data collection. XJ designed statistical analysis plan of the present study, analyzed data, and wrote the manuscript under FX's direction. HG, ML, KS and FX provided critical revisions of the manuscript. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

All procedures performed in this research involving human participants were in accordance with the ethical standards of Peking University and have been performed in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all patients included in the study. Participants can withdraw at any time without any consequences.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main St W, Hamilton, ON L8S 4K1, Canada. <sup>2</sup>China Center for Health Economic Research, Peking University, Beijing 100800, China. <sup>3</sup>National School of Development, Peking University, Beijing 100800, China. <sup>4</sup>Department of Medicine, McMaster University, Hamilton, ON L8S 4K1, Canada. <sup>5</sup>Department of Sociology, McMaster University, Hamilton, ON L8S 4K1, Canada. <sup>6</sup>School of Pharmaceutical Science, Peking University, Beijing 100800, China. <sup>7</sup>School of International Pharmaceutical Business, China Pharmaceutical University, Nanjing 211198, Jiangsu, China. <sup>8</sup>Centre for Evaluation of Medicines, Father Sean O'Sullivan Research Centre, St. Joseph's Healthcare Hamilton, Hamilton L8N 4A6, Canada.

Received: 30 May 2017 Accepted: 18 April 2018

Published online: 27 April 2018

**References**

- International Diabetes Federation. IDF Diabetes Atlas. 7th ed. Brussels: International Diabetes Federation; 2015.
- Isla Pera P. Living with diabetes: quality of care and quality of life. *Patient Prefer Adherence*. 2011;5:65–72.
- Rubin RR. Diabetes and quality of life. *Diabetes Spectrum*. 2000;13:21.
- Rubin RR, Peyrot M. Quality of life and diabetes. *Diabet Metab Res Rev*. 1999;5:205–18.
- Ware JE Jr, Gandek B, Guyer R, Deng N. *Health Qual Life Outcomes*. 2016;14:84–99.
- El Achhab Y, Nejjarri C, Chikri M, Lyoussi B. Disease-specific health-related quality of life instruments among adults diabetic: a systematic review. *Diabetes Res Clin Pract*. 2008;80:171–84.
- Watkins K, Connell CM. Measurement of health-related QOL in diabetes mellitus. *PharmacoEconomics*. 2004;22:1109–26.
- Jacobson AM, Barofsky I, Cleary P, Rand LL. Reliability and validity of a diabetes quality-of-life measure for the diabetes control and complications trial (DCCT). *Diabet Care*. 1988;11:725–32.
- Jacobson AM. Quality of life in patients with diabetes mellitus. *Semin Clin Neuropsychol*. 1997;2:82–93.
- PROQOLID. Diabetes Quality of Life Measure (DQOL). <https://eprovide.mapi-trust.org/instruments/diabetes-quality-of-life-measure#languages>. Accessed 18 Dec 2016.
- Cheng AY, Tsui EY, Hanley AJ, Zinman B. Cultural adaptation of the diabetes quality-of-life measure for Chinese patients. *Diabetes Care*. 1999;22:1216–7.
- Cheng AY, Tsui EY, Hanley AJ, Zinman B. Developing a quality of life measure for Chinese patients with diabetes. *Diabetes Res Clin Pract*. 1999;46:259–67.
- Ding Y, Kong D, Ni Z, Deng H. Culture adaption and revision of diabetes-specific quality of life scale (QDOL). *Chin J Behav Med Sci*. 2004;13:102–3.
- Ding Y, Ni Z, Zhang J, Chen G, Feng H. The assessment on reliability and validity of adjusted diabetes quality of life (A-DQOL) scale. *Chin J Prev Contr Chron Non-commun Dis*. 2000;8:160–2.
- Liang M. Quality of life and its master factors of community-based patients with type 2 diabetes in Beijing. Master degree thesis. Beijing: Beijing university of Chinese medicine; 2014.
- Chen A, Su A, Bai H. Factors impact on the quality of life of type 2 diabetic patients. *Chin Community Doctors*. 2009;11:38.
- Hou Y, Yang Q. Current research status of health-related quality of life in patients with diabetes mellitus. *Chin J Clinicians*. 2016;10:433–6.
- Ren Z. Quality of life in the patients of diabetes mellitus with micro-albuminuria but no renal insufficiency. Master degree thesis. Hangzhou: Zhejiang University; 2010.
- Li D, Ma A, Li H. Systematic review of diabetes-specific quality of life measures in China. *Chin J Pharm Econ*. 2012;34:45–52.
- Qu L, Pan M. Research progress of diabetes-specific quality of life measures in China. *Chin J Behav Med Sci*. 2007;16:765–6.
- Streiner DL, Norman GR, Cairney J. *Health measurement scales. A practical guide to their development and use*. 5th ed. Oxford: Oxford University Press; 2015.
- Cappelleri JC, Lundy JJ, Hays RD. Overview of classical test theory and item response theory for quantitative assessment of items in developing patient-reported outcome measures. *Clin Ther*. 2014;36:648–62.
- ten Holt JC, van Duijn MAJ, Boomsma A. Scale construction and evaluation in practice: a review of factor analysis versus item response theory applications. *Psychol Test Assess Model*. 2010;52:272–97.
- Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*. 2007;16:5–18.
- Weinhardt JM, Morse BJ, Chimeli J, Fisher J. An item response theory and factor analytic examination of two prominent maximizing tendency scales. *Judgm Decis Mak*. 2012;7:644–58.
- The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med*. 1998;28:551–8.
- Prieto L, Alonso J, Lamarca R. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes*. 2003;1:27.
- Fayers PM, Machin D. *Quality of life: the assessment, analysis and reporting of patient-reported outcomes*. 3rd ed. West Sussex: John Wiley & Sons; 2016.
- Kline TJB. *Psychological testing: a practical approach to design and evaluation*. Thousand Oaks, CA: Sage Publications Inc; 2005.
- Norman G. Likert scales, levels of measurement and the “laws” of statistics. *Adv Health Sci Educ*. 2010;15:625–32.
- Sullivan GM, Artino AR. Analyzing and Interpreting Data From Likert-Type Scales. *J Grad Med Educ*. 2013;5:541–2.
- Thompson N. Interpreting item statistics from classical test theory. 2015. <http://www.assess.com/interpreting-item-statistics-from-classical-test-theory/>. Accessed 20 Apr 2018.
- Lester PE, Inman D, Inman Freitas DL, Bishop LK. *Handbook of tests and measurement in education and the social sciences*. Lanham: Rowman & Littlefield Publishing Group; 2014.
- Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw Hill; 1994.
- Qiu H. *Quantitative research and statistical analysis*. Chongqing: Chongqing University Press; 2009.
- Bartlett MS. Tests of significance in factor analysis. *Br J Stat Psychol*. 1950;3:77–85.
- Kaiser HF. A second generation little jiffy. *Psychometrika*. 1970;35:401–15.
- Osborne JW. *Best practices in exploratory factor analysis*. Washington, US: CreateSpace Independent Publishing Platform; 2014.
- Ledesma RD, Valero-Mora P, Macbeth G. The scree test and the number of factors: a dynamic graphics approach. *Span J Psychol*. 2015;18:1–10.
- Osborne JW. What is rotating in exploratory factor analysis? *PARE*. 2015;20
- Thompson B. *Exploratory and confirmatory factor analysis: understanding concepts and applications*. Washington, DC: American Psychological Association; 2004.
- Pallant J. *SPSS survival manual: A step by step guide to data analysis using SPSS for windows*. 3rd ed. New York, USA: Mc Graw Hill; 2007.
- Gliem JA, Gliem R R. Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. 2003. <https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem+&+Gliem.pdf?sequence=1>. Accessed 26 Dec 2016.
- Yamamoto K. Hybrid model of IRT and latent class models. *ETS Research Report Series*. 1982. <https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.2333-8504.1982.tb01326.x>. Accessed 20 Apr 2018.
- Samejima F. Estimation of latent ability using a response pattern of graded scores (psychometric monograph no. 17). Richmond, VA: psychometric Society 1969. <https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf>. Accessed 20 Apr 2018.
- Zickar MJ, Russell SS, Smith CS, Bohle P, Tilley AJ. Evaluating two morningness scales with item response theory. *Pers Individ Dif*. 2002;33:11–24.
- Hooper D, Coughlan J, Mullen MR. Structural equation modelling: guidelines for determining model fit. *J Bus Res Meth*. 2008;6:53–60.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6(1):1–55.
- Liu GG, Wu H, Li M, Gao C, Luo N. Chinese time trade-off values for EQ-5D health states. *Value Health*. 2014;17:597–604.
- EuroQol Group. EQ-5D-3L User Guide. [https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-3L\\_UserGuide\\_2015.pdf](https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-3L_UserGuide_2015.pdf). Accessed 20 Apr 2018.
- Chen X, Qiu Z, Gu M, Su Y, Liu L, Liu Y, Mo C, Xu Q, Sun J, Li D. Translation and validation of the Chinese version of the Quality of Life Radiation Therapy Instrument and the Head & Neck Module (QOL-RTI/H&N). *Health Qual Life Outcomes*. 2014;12:51.
- Global Affairs Canada. Cultural Information – China. [https://www.international.gc.ca/cil-cai/country\\_insights-apercus\\_pays/ci-ic\\_cn.aspx?lang=eng](https://www.international.gc.ca/cil-cai/country_insights-apercus_pays/ci-ic_cn.aspx?lang=eng). Accessed 16 Jan 2017.
- National Bureau of Statistics of China. *China Statistical Yearbook 2013*. 2013. <http://www.stats.gov.cn/tjsj/ndsj/2013/indexch.htm>. Accessed 26 Dec 2016.
- Moss S. Fit indices for structural equation modeling. <http://www.sicotests.com/psyparticle.asp?id=277>. Accessed 26 Dec 2016.

55. Kenny DA, McCoach DB. Effect of the number of variables on measures of fit in structural equation modeling. *Struct Equ Modeling*. 2003;10:333–51.
56. Bogen K. The effect of questionnaire length on response rates—A review of the literature. <https://www.census.gov/srd/papers/pdf/kb9601.pdf>. Accessed 16 Apr 2018.
57. Fan X. Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Educ Psychol Meas*. 1998;58:357–81.
58. StataCorp. Stata Item Response Theory Reference Manual (Release 14). College Station, TX: StataCorp LP. 2015.<http://www.stata.com/manuals14/irt.pdf>. Accessed 26 Dec 2016.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

