

# Comparing the SF-36 and SF-12 in Psychometric Properties as Measuring Quality of Life among Adolescent in China: a Large Sample Cross-sectional Study

**Yanwei Lin**

Guangdong Medical University

**Yulan Yu**

Guangdong Medical University

**Jiayong Zeng**

Guangdong Medical University

**Xudong Zhao**

Tongji University

**Chonghua Wan** (✉ [wanchh1964@qq.com](mailto:wanchh1964@qq.com))

Guangdong Medical University <https://orcid.org/0000-0002-4546-0620>

---

## Research

**Keywords:** Quality of Life; Reliability; validity; Discrimination; Average Information

**Posted Date:** April 3rd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-19671/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published on November 9th, 2020. See the published version at <https://doi.org/10.1186/s12955-020-01605-8>.

# Abstract

**Objective:** By comparing psychometric properties of the SF-36 and the SF-12, supplied evidence for the election of instruments of the quality of life (QOL) and decision-making processes to promote the Quality of Life of adolescent. **Methods:** Stratified cluster random sampling was adopted. The Short-Form 36 (SF-36) was used to assess QOL. Pearson Correlation Coefficient was used to show correlation. Cronbach's Alpha and Construct Reliability (CR) were used to evaluate reliability of SF-36 and the Short-Form 12 (SF-12), Criterion Validity and Average Variance Extracted (AVE, Convergence Validity) for validity. Confirmatory factor analysis was used to calculate load factor for each item, then obtained CR and AVE. The Semejima grade response model (Logistic two-parameter module) in the item response theory was used to estimate the Item Discrimination, Item Difficulty and Item Average Information of each item. **Results:** 19,428 samples were included in the study. The mean age was 14.78 years (SD=1.77). High correlations between corresponding domains and components of both scales were found. Reliability of sf-36 each domain was better than that corresponding domain of sf-12. Domains of PF, RP, BP, and GH in SF-36 had good construct reliability (CR,>0.6). The Criterion Validities of SF-36 were little higher in some corresponding dimensions except PCS. Convergence validities of SF-12 were higher than SF-36 in PF, RP, BP and PCS. The items of BP, SF, RP and VT in SF-12 had acceptable discriminations of items and higher than in SF-36. The items Average Amounts of Information of BP, VT, SF, RE and MH in SF-36 and SF-12 were poor. **Conclusion:** Two components (PCS and MCS) measurements of SF-12 appeared to perform at least as well as the SF-36 in cross-sectional settings in adolescence. Some domains, for instance SF and BP, were suitable for adolescents or not need study further.

## 1. Introduction

Youth involved identity building; such experiences could shape their attributes and attitudes, leading to risky behaviors in their lives.[1] Due to individual experiences experiments and transformations, the determinants of health and disease for adolescence traversed the social and psychological fields [2]. A deeper understanding of how adolescents view their lives allowed a greater understanding of their health. Health-related quality of life of school adolescents in some international studies was discussed. 'Health-related quality of life' (HRQOL) was a comprehensive model of subjective health, which had covered physical, social, psychological and functional aspects of individual well-being as a multidimensional and subjective construct [3, 4]. For the purpose of guiding the organization of resources and decision-making processes to promote the quality of life of adolescent, Understanding the quality of adolescent's life was essential[5, 6].

the Short-Form 36(SF-36)was developed and validated as the most appropriate instrument to generic short form health survey for measuring Quality of Life (QOL), which was widely applied to assess important QOL domains in the Medical Outcomes Study[7]. The SF-36 consists of eight QOL domains (PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health) that comprise two summary measures-the physical component summary (PCS, calculated from PF RP, BP, and GH) and the mental component summary (MCS ,calculated from VT, SF, RE, and MH)[8]. One of the major advantages of using the SF-36 is that it allows for QOL scores to be compared to scores in different groups[9], However, because the SF-36 was not originally designed to measure important QOL domains specific to adolescent, some studies presented the SF-36, especially the mental component summary, to be relatively insensitive to variations in different populations over time[10–12].

A substantially shorter questionnaire, the SF-12 that was developed by Ware and colleagues utilized a reduced number of items from 36 to 12 for reducing the considerable burden placed on respondents and investigators generically by SF-36 [13, 14]. Most of respondents completed the SF-12 in less than a third of the usual time needed to complete the SF-36 [8]. Ware showed the two instruments highly correlated, and about 90% of the variation in both of the physical and mental component summary measures in the SF-36 was explained by the same summary measures of the SF-12[15]. Subsequent studies that compared the two scales had suggested varying results on account of the disease or health condition of interest. [16–18]

The SF-12 and SF-36 were available in many languages, and were applied to all kinds of groups, including in adolescence[19]. Although studies had demonstrated that both scales were valid instruments for adolescent, they were rarely used to evaluate QOL of adolescent in china, In other words, few studies had focused on the quality of life of healthy adolescents in china[2, 20].

In adolescence, studies surveying the perception of QOL in chronic patients that conducted in hospital or outpatient settings were predominant [21, 22]. Otherwise, a recent interest in the study of healthy groups had accreted and been performed in other contexts, such as in school[23, 24], because it was beneficial to recognizing and monitoring of adolescents vulnerable to a poor health-related quality of life[25, 26]. In some studies, though the SF-12 and SF-36 were used to investigate to the perception of QOL in adolescent, It was unclear which of the two scales was more suitable to adolescent[27].

Thus, our study aimed to evaluate the QOL of adolescent students at school in china by using the SF-36 and SF-12, through comparing psychometric properties of the SF-36 and the SF-12, supplying evidence for the election of instruments of the quality of life and decision-making processes to promote the quality of life of adolescent.

## 2. Methods

### 2.1 Study design and Sample

Stratified cluster random sampling was adopted[28], firstly, dividing regions by geographical location, and Guangdong, Shanghai, Shenyang, Wuhan, Xi'an and Yunnan represented the south, east, north, central, northwest and southwest regions respectively. These areas were chosen in order to ensure proper representation by including participants from geographically diverse areas. Secondly, middle schools were randomly selected and followed by grade (First grade of junior school to third grade of high school), and all of students enrolled and effectively attending in the selected classes were eligible, except for those with any physical or mental condition that cannot complete questionnaires. The study was approved by the Institutional Review Board (IRB) at Affiliated Hospital of Guangdong Medical University. Verbal informed consent was obtained for publication from the participants and/or their relatives as approved by the IRB. The response rate was almost 80%. This present study included the 19428 adolescents with complete information on quality of life measures. The sample sizes for each region were Guangdong (4490, 23.1%), Shanghai (1039, 5.3%), Shenyang (3539, 18.2%), Wuhan (1371, 7.1%), Xi'an (4197, 21.6%) and Yunnan (4792, 24.7%)

### 2.2 Instruments and variable

SF-36 was used to assess QOL. Comprising eight subscales-physical functioning (PF), role functioning (RF), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role of emotional (RE) and mental health (MH), the first four subscales constituted the physical component summary (PCS-36) among them, the remaining four subscales made up the mental component summary (MCS-36). Based on the response to individual items comprising that subscale and using a z-score transformation, Scores of each subscale are calculated. Using standard methods, aggregated to estimate physical and mental summary scores[29].

SF-12 component summary scores (eight subscales, PCS-12 and MCS-12) were calculated using SF-12 items embedded in the SF-36[30]. It had been presented to be equivalent to calculating SF-12 derived from the SF-12 as a standalone questionnaire[16]. All summary scores range from 0–100 where higher scores indicated better QOL.

### 2.3 Statistical analysis

For descriptive analyses, we aimed to show overall demographics, and QOL. We calculated average and standard deviations in QOL scores by SF-36 and SF-12. For testing the relevance of them, Pearson Correlation Coefficient was used to show correlation between subscales of SF-36 and SF-12.

Cronbach's Alpha and Construct Reliability (CR) were used to evaluate reliability of SF-36 and SF-12, and validity indicators were represented by criterion validity and average variance extracted (AVE). Confirmatory factor analysis was used to calculate load factor for each item, then obtained CR and AVE according to load factors. Criterion validity was expressed by the correlation between the response of each subscale and self-reported health status.

According to the evaluation results of the samples, and taking into account the characteristics of the ordered and multi-category forms of the scale items, the Semejima grade response model (Logistic two-parameter module) in the item response theory was used to estimate the item discrimination, item difficulty and item average information of each item[31]. Multilog 7.03 and Amos 20.0 were used to process data.

## 3. Results

### 3.1 Sample characteristics

Of the 20,226 questionnaires received, 798 had no responses on some the SF-36 items. Finally, 19,428 samples were included in the study. The mean age of the sample of respondents was 14.78 years (standard deviation; SD = 1.77), 49.4% (9595) were boys. Among the sf-36 and sf-12 scales, the physical functioning (PF) mean score was the highest, and the role emotional (RE) mean scores was the lowest. The biggest mean difference in scores between the two scales was in the social functioning domain (SF). Corresponding domains of two scales, the role emotional dimension was the most relevant ( $r = 0.923$ ). Details in Table 1.

Table 1  
Scores of SF-36, SF-12 and SF-8 among adolescent (n = 19428)

	SF-36	SF-12	Mean difference	Correlation coefficient
PF***	89.10 ± 14.39	91.64 ± 16.85	-2.54	0.800
RP***	68.86 ± 34.28	68.08 ± 39.44	0.78	0.897
BP***	79.97 ± 19.77	85.09 ± 19.25	-5.12	0.876
GH***	70.41 ± 19.53	62.72 ± 26.39	7.69	0.670
VT***	65.04 ± 17.19	62.11 ± 25.90	2.93	0.645
SF***	77.98 ± 19.07	66.17 ± 23.17	11.81	0.875
RE***	54.82 ± 37.45	52.14 ± 40.44	2.68	0.923
MH***	68.51 ± 17.18	64.86 ± 18.83	3.65	0.799
PCS***	75.00 ± 11.10	70.52 ± 13.65	4.48	0.812
MCS***	68.55 ± 14.18	61.32 ± 7.17	7.23	0.779

Abbreviations: PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health; PCS, physical component summary; MCS, mental component summary. PCS was calculated from PF, RP, BP, and GH, and that for MCS was calculated from VT, SF, RE, and MH. \*p < 0.05; \*\* p < 0.01; \*\*\*p < 0.001

### 3.2 Psychometric properties in classical test theory

Except for SF, scales composed of multiple items had generally acceptable internal reliability (Table 2). The low internal reliability of SF was probably because of inconsistent understanding of the meaning of the corresponding items, the only two items (one what extent has your physical health or emotional problems interfered with and one how much of the time has your physical health or emotional problems interfered with) of SF may be biased in understanding for adolescents. Moreover, consistent with related studies, internal reliability of MH of SF-12 was low (0.369). On the other hand, internal reliability of sf-36 each domain was better than that corresponding domain of sf-12, which was consistent with higher internal reliability for more items. Domains of PF, RP, BP, and GH in SF-36 had good construct reliability (CR,>0.6), so, PCS made up of PF, RP, BP, and GH had high construct reliability too. Except for RP and PCS, domains in SF-12 were not good at construct reliability.

Criterion Validity was calculated based on the item of self –reported health (In general, would you say your health is). It is worth noting that Criterion Validities of domains of the two scales were low, especially in the domains of PF, RP and SF, which suggested that correlation between physical health and self-perceived health was weak. Moreover, in PCS, Criterion Validity of SF-12 was much higher than the Criterion Validity of SF-36. Although the Criterion Validities of SF-36 were higher in other corresponding dimensions, the differences were little. Average Variance Extracted (AVE) was also called Convergence Validity, PF, RP, BP and PCS had generally acceptable convergence validity Whether SF-36 or SF-12, moreover, in these domains, convergence validities of SF-12 were higher than SF-36, and little difference in other domains (Table 2). The load factor results for confirmatory factor analysis that were used to calculate CR and AVE were shown in Fig. 1.

Table 2  
Validity and reliability of SF-36 and SF-12 in classical test theory

	SF-36				SF-12				Difference(SF-36- SF-12)			
	reliability		validity		reliability		validity		reliability		validity	
	Cronbach's Alpha	CR	Criterion Validity	AVE	Cronbach's Alpha	CR	Criterion Validity	AVE	Cronbach's Alpha	CR	Criterion Validity	AVE
PF	0.841	0.860	0.085	0.389	0.564	0.581	0.055	0.410	0.277	0.279	0.030	-0.021
RP	0.727	0.730	0.173	0.405	0.605	0.620	0.171	0.449	0.122	0.110	0.002	-0.044
BP	0.670	0.690	0.283	0.528	-	0.593	0.227	0.593	-	0.097	0.056	-0.065
GH	0.766	0.781	0.670	0.420	-	0.292	-	0.292	-	0.489	-	0.128
VT	0.569	0.577	0.309	0.302	-	0.068	0.252	0.068	-	0.509	0.057	0.234
SF	0.211	0.329	0.113	0.146	-	0.203	0.027	0.203	-	0.126	0.086	-0.057
RE	0.626	0.489	0.203	0.371	0.485	0.460	0.199	0.331	0.141	0.029	0.004	0.04
MH	0.625	0.426	0.243	0.316	0.396	0.398	0.049	0.313	0.229	0.028	0.194	0.003
PCS	0.562	0.935	0.350	0.430	0.422	0.820	0.589	0.434	0.14	0.115	-0.239	-0.004
MCS	0.609	0.418	0.476	0.299	0.429	0.383	0.300	0.260	0.18	0.035	0.176	0.039
Construct Reliability = CR, Average Variance Extracted = AVE												

Figure 1 standardized parameter estimates for confirmatory factor analysis of the Short Form-36

## 2.3 Psychometric properties in item response theory

According to the Samezima grade response model, the parameter values and information content of the items were shown in Table 3. It could be seen that the discriminations of items were between 0.45–2.73, with a large gap. The difficulties of items were ascending from the lowest level to the highest level unidirectionally, which meet the difficulty assumptions estimated by the model. Average amount of information of each item was between 0.07–1.02.

In SF-36, the domains of PF, RP, GH and RE had acceptable discriminations of items (> 1), the remaining dimensions were less differentiated, especially BP and SF, probably because for teenagers, there was strong homogeneity between individuals in terms of physical pain and social function. On the other hand, in SF-12, BP, SF, RP and VT had higher discriminations of items than in SF-36.

With reference to relevant literatures, the amount of information measured on the scale > 25 indicated that the quality of the evaluation items was good; the amount of information < 16 indicates that the evaluation items were poor. Combining the number of items on the scale, for SF-36, divided 16 and 25 by 36, respectively, to get the average information amount of each item, so as to obtain the determination criterion: the average information amount of the item > 0.69 (25/36) For excellent, items < 0.44 (16/36) were judged to be poor. Similarly, for SF-12, the average information amounts of the items > 2.08 were judged to be excellent, and items < 1.33 were judged to be poor. Except for PF05 and PF09, the items of PF domain of SF-36 were excellent, the items of GH domain of SF-36 were excellent too, but, the items of BP, VT, SF, RE and MH were poor in SF-36. On the other hand, Average Amounts of Information of SF-12 items were poor.

Table 3  
item discrimination, difficulty and average amount of information in item response theory

Label	SF-36			SF-12		
	Item Discrimination (SD)	Item Difficulty(SD)	Average Amount of Information	Item Discrimination (SD)	Item Difficulty (SD)	Average Amount of Information
physical functioning(PF)						
PF01	2.73(0.01)	-1.43(0.01),0.21(0.01)	1.02			
PF02	2.73(0.01)	-2.53(0.05),-1.07(0.01)	0.74	2.20(0.03)	-3.13(0.05) -1.40(0.02)	0.45
PF03	2.73(0.01)	-2.55(0.05),-1.14(0.01)	0.73			
PF04	2.73(0.01)	-2.05(0.03),-0.87(0.01)	0.90	2.20(0.03)	-2.60(0.04) -1.17(0.01)	0.54
PF05	2.73(0.01)	-2.45(0.04),-1.54(0.02)	0.65			
PF06	2.73(0.01)	-1.88(0.02),-0.90(0.01)	0.89			
PF07	2.73(0.01)	-1.42(0.01),-0.25(0.01)	0.95			
PF08	2.73(0.01)	-1.96(0.03),-0.92(0.01)	0.89			
PF09	2.73(0.01)	-2.51(0.05),-1.58(0.02)	0.63			
PF10	2.73(0.01)	-1.69(0.02),-1.20(0.01)	0.74			
role physical(RP)						
RP1	2.17(0.02)	0.77(0.01)	0.43			
RP2	2.17(0.02)	0.53(0.01)	0.43	2.32(0.03)	0.52(0.01)	0.46
RP3	2.17(0.02)	0.65(0.01)	0.43	2.32(0.03)	0.63(0.01)	0.46
RP4	2.17(0.02)	0.52(0.01)	0.43			
bodily pain(BP)						
BP1	0.45(0.01)	-10.26(0.48),-8.08(0.31),-4.60(0.16) -1.33(0.06),1.24(0.06)	0.06			
BP2	0.45(0.01)	0.32(0.05),4.65(0.17),7.46(0.28) 10.00(0.44)	0.05	1.06(0.02)	0.18(0.02) 2.40(0.04) 3.83(0.07) 5.28(0.12)	0.24
general health(GH)						
GH1	1.76(0.01)	-3.05(0.05),-1.11(0.02),-0.13(0.01) 1.2(0.01)	0.76	0.91(0.01)	-1.80(0.03) 0.21(0.02) 1.72(0.03) 4.93(0.09)	0.24
GH2	1.76(0.01)	-2.33(0.03),-1.46(0.02),-0.23(0.01) 0.54(0.01)	0.73			
GH3	1.76(0.01)	-2.77(0.03),-2.14(0.02),-0.89(0.01) 0.35(0.01)	0.68			
GH4	1.76(0.01)	-2.43(0.03),-1.55(0.02),-0.52(0.01) 0.17(0.01)	0.67			
GH5	1.76(0.01)	-2.75(0.03),-2.04(0.02),-0.78(0.01) 0.57(0.01)	0.71			
Vitality(VT)						
VT1	0.74(0.00)	-2.43(0.04),0.29(0.03),1.68(0.03) 3.33(0.05),4.71(0.07)	0.17			
SD = standard deviation						

Label	SF-36			SF-12		
VT2	0.74(0.00)	-2.74(0.04),-0.40(0.03),1.22(0.03) 2.89(0.04)	0.17	0.91(0.01)	-2.36(0.01) -0.35(0.02) 1.07(0.02) 2.50(0.04) 3.90(0.07)	0.25
VT3	0.74(0.00)	-4.73(0.07),-3.10(0.04),-1.97(0.03) -0.50(0.03),2.10(0.03)	0.16			
VT4	0.74(0.00)	-4.26(0.06),-2.55(0.04),-1.36(0.03) 0.15(0.03),2.93(0.04)	0.18			
social functioning(SF)						
SF1	0.50(0.01)	-1.68(0.06),2.80(0.08),5.92(0.17) 8.66(0.29)	0.07			
SF2	0.50(0.01)	-6.35(0.18),-4.73(0.13),-3.48(0.10) -2.06(0.07),-0.01(0.05)	0.07	1.07(0.02)	-3.42(0.06) -2.58(0.04) -1.92(0.03) -1.15(0.02) -0.02(0.02)	0.28
role emotional(RE)						
RE1	1.82(0.02)	0.35(0.01)	0.36			
RE2	1.82(0.02)	0.23 (0.01)	0.36	1.63(0.02)	0.24 (0.01)	0.31
RE3	1.82(0.02)	-0.07 (0.01)	0.36	1.63(0.02)	-0.07(0.01)	0.32
mental health(MH)						
MH1	0.78(0.00)	-4.35(0.07),-2.59(0.04),-1.53(0.03) -0.33(0.03),1.50(0.03)	0.19			
MH2	0.78(0.00)	-4.49(0.07),-2.99(0.04),-2.12(0.03) -1.01(0.03),0.82(0.03)	0.18			
MH3	0.78(0.00)	-11(-),-2.84(0.07),-0.42(0.03) 1.03(0.03),2.72(0.04)	0.19	0.79(0.01)	-3.94(0.06) -2.24(0.03) -0.82(0.03) 0.55(0.03) 2.91(0.04)	0.20
MH4	0.78(0.00)	-4.83(0.08),-3.18(0.05),-2.15(0.03) -0.82(0.03),2.04(0.03)	0.18	0.79(0.01)	-4.73(0.07) -3.18(0.04) -2.19(0.03) -0.88(0.03) 2.00(0.03)	0.18
MH5	0.78(0.00)	-11.18(-),-1.35(0.04),0.84(0.03) 2.05(0.03),3.48(0.05)	0.18			
HT	0.91(0.00)	-4.84(0.09),-2.59(0.04),-0.23(0.02) 1.25(0.03)	0.23			
SD = standard deviation						

## 4. Discussion

Psychometric standards were used to evaluate reliability and validity of the standard Chinese SF-36 and SF-12 scales in a large sample of Chinese adolescents in our study. Our study suggested that the SF-12 and SF-36 correlated very highly in Chinese adolescents. Although the reliability and average amount of information of the SF-12 domain and item were lower than that of SF-36, the convergence validity and item discrimination were better partly. No matter SF-36 and SF-12, Psychometric properties of two components (PCS and MCS) were better than the domains.

Studies showed the two scales discriminated between adolescents with physical and mental health problems and performed well in associating with other clinical criteria [19, 32, 33]. A study of 31,357 adolescents in Hong Kong showed the two components and a single general health component of the standard Chinese SF-12 were appropriate health indicators for Chinese adolescents [20]. Studies have also shown that the SF-12 correlated highly with SF-36 in obese and non-obese patients [3, 4]. However, many problems still existed, such as high

correlation between two components, low internal reliability and ceiling effect of individual domain[34]. Comparing the SF-12 and SF-36, previous studies in patients with specific diseases or health conditions have generally found moderate to high correlations between corresponding domains and components of both scales[18, 35]. Our study also demonstrated these correlations. Since the SF-12 is embedded in the SF-36, we expected reasonably high correlations. Overall, the dimensions of the SF-12 scale could reflect 64.5–92.3% of the corresponding dimensions of the SF-36 scale.

A low reliability and validity of social functioning domain was also noted. This might indicate questionable reliability and validity of the instruments or the lack of representation[3]. On the other hand, also it be attributed to the presence of inconsistent responding, which might occur when respondents completed a questionnaire without comprehending the items in adolescents[20]. Due to the brevity of the SF-12 instrument, related research showed it was not possible to get reliable information for each of the eight domains of SF-12 so that one would not be able to draw conclusions about specific domains[36]. Indeed, we found SF-36 was better than SF-12 in reliability. At the same time, compared with SF-12 and SF-36 in validity, no loss in effectiveness had been shown, even a slight improvement. But, we also found criterion validities of PF, SF and MH were low, self-reported health was our criterion. relevant research found that adolescents performed moderate activities or climbing several flights of stairs would not present problems for most because of typically physically fit and active, combining limited social life and adolescent mental state, inconsistent responding would be possible[20].

Unlike previous research[34, 36–38], we found domains of BP and SF had poor discriminations of items, instead of PF, and BP, SF, RP and VT of SF-12 had higher discriminations of items than in SF-36. We thought Compared with PF items, the items in other domains were not easy for teenagers to understand, resulting in a lack of sensitivity in the measurement of adolescents. Similarly, a loss of information had been found in SF-12 that would be provided by the eight dimensions of the SF-36, but, utilization of the two summary dimensions of SF-12 had the advantage based on adolescents, which was consistent with the results of other population studies[20].

Methodological limitations should be mentioned. The participants were stratified regarding geographical areas for minimizing the risk of possible regional. However, the regions chosen were vast and concluded small towns and big cities as well as rural areas[39, 40]. Differences due to these circumstances might exist, but not come to light in this design. Additionally, there was a difference in response consistency between samples because of characteristics of adolescence, leading to bias in results[41].

## 5. Conclusion

In general, our study suggested that the SF-12 correlated highly with the SF-36 in adolescence groups in china, If you only focused on two components (PCS and MCS) measurements, and SF-12 appeared to perform at least as well as the SF-36 in cross-sectional settings in adolescence; hence, using the SF-12 in place of the SF-36 might be appropriate in this situation. Simultaneously, whether some domains, for instance SF and BP, were suitable for adolescents need study further.

## Abbreviation

QOL, quality of life; SF-36, The Short-Form 36; SF-12, The Short-Form 12; PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health; PCS, physical component summary; MCS, mental component summary; CR, construct reliability ; AVE, average variance extracted.

## Declarations

### Availability of data and materials

The study data is available upon request.

### Acknowledgments

We appreciate all participants and the schools involved in the survey, as well as other staff members on the scene.

### Funding

This study was Supported by National Natural Science Foundation of China (Grant number: 30860248, Grant number: 71804029), National Key Technologies Research and Development Program of China (Grant number: 2009BAI77B05), Guangdong Medical Research Foundation (Grant number: C2018081) and Doctoral research start-up foundation of Guangdong Medical University in 2019 (B2019033).

### Author Contributions



Conceived and designed the study: C.W. and X. Z. Performed the study: Y. L., Y. Y. and J. Z., Analyzed the data: Y. L., Wrote the paper: Y. L., All authors have read and approved the manuscript.

### Author details

<sup>1</sup>Department of Health Sociology, School of Humanities and Management, Guangdong Medical University, 1#, Xincheng Avenue, Songshanhu District, Dongguan, 523808, Guangdong, China.

<sup>2</sup>Department of psychology, School of Humanities and Management, Guangdong Medical University, 1#, Xincheng Avenue, Songshanhu District, Dongguan, 523808, Guangdong, China.

<sup>3</sup>Institute of Psychosomatic Medicine, the East Translational Medicine Platform of Tongji University, 50#, Chifeng Avenue, Shanghai, 200092, China.

<sup>4</sup>School of Humanities and Management, Research Center for Quality of Life and Applied Psychology, Guangdong Medical University, 1#, Xincheng Avenue, Songshanhu District, Dongguan, 523808, Guangdong, China.

\* Correspondence: [zhaoxd@tongji.edu.cn](mailto:zhaoxd@tongji.edu.cn) (X. Z.); [wanchh1964@qq.com](mailto:wanchh1964@qq.com) (C. W.)

### Ethics approval and consent to participate

The study was approved by the Institutional Review Board (IRB) at Affiliated Hospital of Guangdong Medical University. Verbal informed consent was obtained for publication from the participants and/or their relatives as approved by the IRB.

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests.

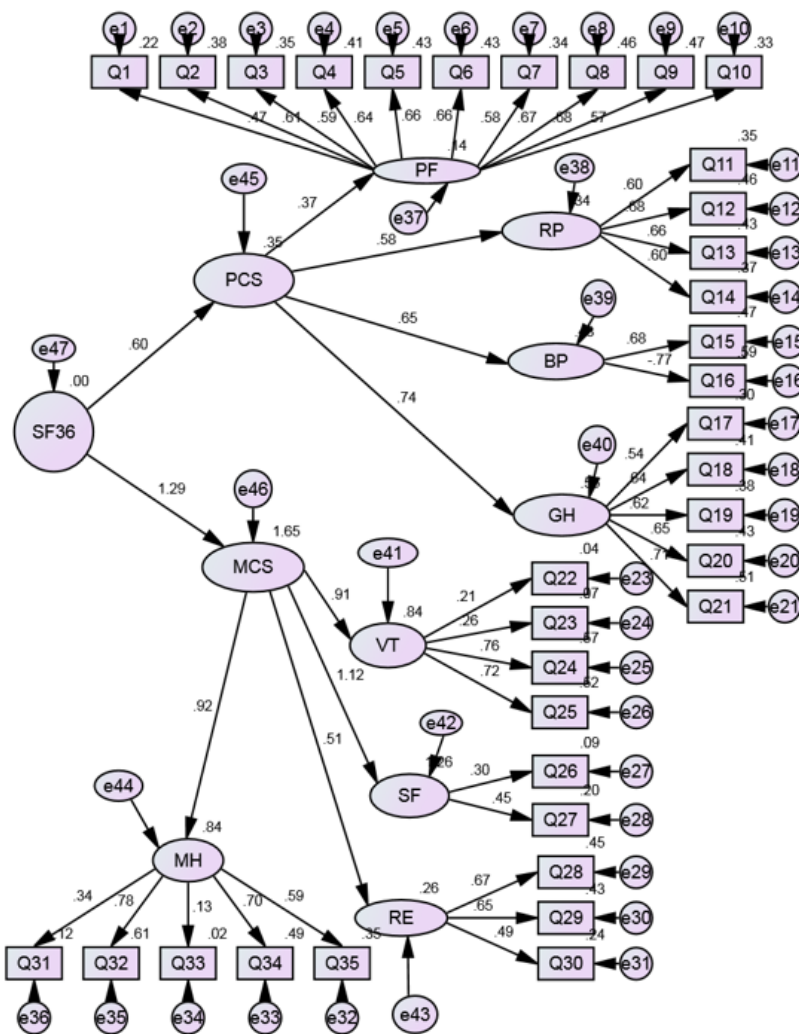
## References

1. Goodall C, Barnard A: Approaches to Working with Children and Families: A Review of the Evidence for Practice. *Practice* 2015, 27:335-351.
2. Agathao BT, Reichenheim ME, Moraes CL: Health-related quality of life of adolescent students. *Cien Saude Colet* 2018, 23:659-668.
3. Wee CC, Davis RB, Hamel MB: Comparing the SF-12 and SF-36 health status questionnaires in patients with and without obesity. *Health Qual Life Outcomes* 2008, 6:11.
4. Corica F, Corsonello A, Apolone G, Lucchetti M, Melchionda N, Marchesini G: Construct validity of the Short Form-36 Health Survey and its relationship with BMI in obese outpatients. *Obesity (Silver Spring)* 2006, 14:1429-1437.
5. Solans M, Pane S, Estrada MD, Serra-Sutton V, Berra S, Herdman M, Alonso J, Rajmil L: Health-Related Quality of Life Measurement in Children and Adolescents: A Systematic Review of Generic and Disease-Specific Instruments. *Value in Health* 2010, 11:742-764.
6. Ravens-Sieberer U, Devine J, Bevans K, Riley AW, Moon J, Salsman JM, Forrest CB: Subjective well-being measures for children were developed within the PROMIS project: presentation of first results. *J Clin Epidemiol* 2014, 67:207-218.
7. Yang F, Wong CKH, Luo N, Piercy J, Jackson J: Mapping the kidney disease quality of life 36-item short form survey (KDQOL-36) to the EQ-5D-3L and the EQ-5D-5L in patients undergoing dialysis. *The European Journal of Health Economics* 2019.
8. Li J, Zhong D, Ye J, He M, Zhang S-I: Rehabilitation for balance impairment in patients after stroke: a protocol of a systematic review and network meta-analysis. *BMJ Open* 2019.
9. Lam CLK, Tse EYY, Gandek B, Fong DYT: The SF-36 summary scales were valid, reliable, and equivalent in a Chinese population. 58:0-822.
10. Brazier J, Roberts J, Deverill M: The Estimation of a Preference-Based Measure of Health from The SF-36. *Journal of Health Economics* 2002, 21:271-292.
11. Fukuhara S: Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey. *Journal of Clinical Epidemiology* 1998, 51.
12. Escobar A, Quintana JM, Bilbao A, Aróstegui I, Vidaurreta I: Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 2007, 15:273-280.

13. Windsor TD, Rodgers B, Butterworth P, Anstey KJ, Jorm AF: Measuring Physical and Mental Health using the SF-12: Implications for Community Surveys of Mental Health. *Australian & New Zealand Journal of Psychiatry* 2006, 40:797-803.
14. Tucker G, Adams R, Wilson D: New Australian population scoring coefficients for the old version of the SF-36 and SF-12 health status questionnaires. *Quality of Life Research*, 19:1069-1076.
15. Ware J, Jr., Kosinski M, Keller SD: A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996, 34:220-233.
16. Muller-Nordhorn J, Roll S, Willich SN: Comparison of the short form (SF)-12 health status instrument with the SF-36 in patients with coronary heart disease. *Heart* 2004, 90:523-527.
17. Jenkinson C, Layte R, Jenkinson D, Lawrence K, Petersen S, Paice C, Stradling J: A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies? *J Public Health Med* 1997, 19:179-186.
18. Hurst NP, Ruta DA, Kind P: Comparison of the MOS short form-12 (SF12) health status questionnaire with the SF36 in patients with rheumatoid arthritis. *Br J Rheumatol* 1998, 37:862-869.
19. Lacson E, Xu J, Lin SF, Dean SG, Lazarus JM, Hakim RM: A Comparison of SF-36 and SF-12 Composite Scores and Subsequent Hospitalization and Mortality Risks in Long-Term Dialysis Patients. 2009, 5:252.
20. Fong DY, Lam CL, Mak KK, Lo WS, Lai YK, Ho SY, Lam TH: The Short Form-12 Health Survey was a valid instrument in Chinese adolescents. *J Clin Epidemiol* 2010, 63:1020-1029.
21. Sato S, Nishimura K, Tsukino M, Oga T, Hajiro T, Ikeda A, Mishima M: Possible Maximal Change in the SF-36 of Outpatients with Chronic Obstructive Pulmonary Disease and Asthma. *Journal of Asthma*, 41:355-365.
22. Asarnow JR, Jaycox LH, Duan N, LaBorde AP, Rea MM, Murray P, Anderson M, Landon C, Tang L, Wells KB: Effectiveness of a Quality Improvement Intervention for Adolescent Depression in Primary Care Clinics. *Jama the Journal of the American Medical Association*, 293:311.
23. Harding L: Children's Quality of Life Assessments: A Review of Generic and Health Related Quality of Life Measures completed by Children and Adolescents. *Clinical Psychology & Psychotherapy* 2001, 8:79-96.
24. Kontodimopoulos N, Damianou K, Stamatopoulou E, Kalampokis A, Loukos I: Children's and parents' perspectives of health-related quality of life in newly diagnosed adolescent idiopathic scoliosis. *Journal of Orthopaedics* 2018.
25. Paltzer J, Barker E, Witt WP: Measuring the health-related quality of life (HRQoL) of young children in resource-limited settings: a review of existing measures. 22:1177-1187.
26. Spencer N: Socioeconomic determinants of health related quality of life in childhood and adolescence: results from a European study. *Child Care Health & Development* 2006, 32:603-604.
27. Song B, Hu W, Hu W, Yang R, Li D, Guo C, Xia Z, Hu J, Tao F, Fang J, Zhang S: Physical Disorders are Associated with Health Risk Behaviors in Chinese Adolescents: A Latent Class Analysis. *International Journal of Environmental Research and Public Health* 2020, 17:2139.
28. Tipton, E.: Stratified Sampling Using Cluster Analysis: A Sample Selection Strategy for Improved Generalizations From Experiments. *Evaluation Review* 2013, 37:109-139.
29. Gandek B, Jr. JEW, Aaronson NK, Alonso J, Apolone G, Bjorner J, Brazier J, Bullinger M, Fukuhara S, Kaasa S: Tests of Data Quality, Scaling Assumptions, and Reliability of the SF-36 in Eleven Countries: Results from the IQOLA Project. *Journal of Clinical Epidemiology* 1998, 51:0-1158.
30. Gandek B, Ware JE, Jr, Aaronson NK, Apolone G, Sullivan M: Cross-Validation of Item Selection and Scoring for the SF-12 Health Survey in Nine Countries: Results from the IQOLA Project. *Journal of Clinical Epidemiology* 1998, 51:1171-1178.
31. Dimitris R: Irm: An R Package for Latent Variable Modeling and Item Response Analysis. *Journal of Statistical Software* 2006, 17.
32. Failde I, Medina P, Ramirez C, Arana R: Assessing health-related quality of life among coronary patients: SF-36 vs SF-12. *Public Health* 2009, 123:615-617.
33. Waal JMVD, Terwee CB, Windt DIAWMvd, Bouter LM, Dekker J: The Impact of Non-Traumatic Hip and Knee Disorders on Health-Related Quality of Life as Measured with the SF-36 or SF-12. A Systematic Review. *Quality of Life Research* 2005, 14:1141-1155.
34. Nortvedt MW, Riise T, Myhr KM, Nyland HI: Performance of the SF-36, SF-12, and RAND-36 Summary Scales in a Multiple Sclerosis Population. *Medical Care* 2000, 38:1022-1028.
35. Tucker G, Adams R, Wilson D: New Australian population scoring coefficients for the old version of the SF-36 and SF-12 health status questionnaires. *Qual Life Res* 2010, 19:1069-1076.
36. White MK, Maher SM, Rizio AA, Bjorner JB: A meta-analytic review of measurement equivalence study findings of the SF-36® and SF-12® Health Surveys across electronic modes compared to paper administration. *Quality of Life Research An International Journal of Quality of*

37. Huang IC, Wu AW, Frangakis C: Do the SF-36 and WHOQOL-BREF measure the same constructs? Evidence from the Taiwan population\*. Qual Life Res 2006, 15:15-24.
38. Conner-Spady BL, Marshall DA, Bohm E, Dunbar MJ, Noseworthy TW: Comparing the validity and responsiveness of the EQ-5D-5L to the Oxford hip and knee scores and SF-12 in osteoarthritis patients 1 year following total joint replacement. Quality of Life Research An International Journal of Quality of Life Aspects of Treatment Care & Rehabilitation 2018, 27:1-12.
39. Amalraj VA, Balakrishnan R, Jebadhas AW, Balasundaram N: Constituting a Core Collection of Saccharum spontaneum L. and Comparison of Three Stratified Random Sampling Procedures. Genetic Resources & Crop Evolution 2010, 53:1563-1572.
40. Buddhakulsomsiri J, Parthanadee P: Stratified random sampling for estimating billing accuracy in health care systems. Health Care Management Science 2008, 11:41-54.
41. Saigal S: Self-perceived Health Status and Health-Related Quality of Life of Extremely Low-Birth-Weight Infants at Adolescence. Jama 1996, 276:453.

## Figures



**Figure 1**  
standardized parameter estimates for confirmatory factor analysis of the Short Form-36