Journal of Evaluation in Clinical Practice ISSN 1356-1294

Examining the hierarchical factor structure of the SF-36 Taiwan version by exploratory and confirmatory factor analysis

Chia-huei Wu,[1] Keng-lin Lee[2] and Grace Yao[3]

[1]MS Student, [2]PhD Student, [3]PhD Associate Professor, Department of Psychology, National Taiwan University, Taipei, Taiwan

Keywords
confirmatory factor analysis, exploratory factor analysis, health status, SF-36

Correspondence
Grace Yao
Department of Psychology
National Taiwan University
1, Sec. 4, Roosevelt Road
Taipei
Taiwan
E-mail: kaiping@ntu.edu.tw

Accepted for publication: 10 May 2006

doi:10.1111/j.1365-2753.2006.00767.x

Abstract

Rationale The purpose of this study was to examine the factor structure of the Medical Outcome Study Short Form-36 Taiwan version (SF-36 Taiwan version) using data from the 2001 National Health Interview Survey in Taiwan.
Method The 2001 National Health Interview Survey was conducted by stratified multi-stage systematic sampling, resulting in 19 777 valid responses for the SF-36 Taiwan version. In this study, the 19 777 participants were randomly divided into two independent samples. One sample ($n = 9856$) was used for exploratory factor analysis (EFA), and the other ($n = 9921$) was used for confirmatory factor analysis (CFA).
Results The EFA suggested a seven-first-order-factor structure for the SF-36 Taiwan version. In addition, hierarchical EFA revealed that there was only one second-order factor underlying the seven first-order factors. Further, CFA was conducted on the other sample to compare the performances of the original model with eight first-order factors and two second-order factors, and the revised model with seven first-order factors and one second-order factor. The CFA results revealed that the original model was better than the revised model.
Conclusion According to the EFA and CFA, it can be concluded that the original structure is still acceptable for the SF-36 Taiwan version.

Introduction

The Medical Outcome Study Short Form-36 (SF-36) is a widely used instrument for measuring health status. It contains 35 items for eight subscales, including (1) physical functioning (PF), (2) role limitation due to physical problems (RP), (3) bodily pain (BP), (4) general health (GH), (5) vitality (VT), (6) social functioning (SF), (7) role limitation due to emotional problems (RE), and (8) mental health (MH). There is also a single item about perception change over the past 12 months (out of the eight subscales). It has been proposed that there are eight first-order factors (for eight subscales) and two second-order factors (physical and mental health) underlying the SF-36 [1]. Therefore, two summary scores for physical and mental health usually have been used to indicate an individual's health status.

The SF-36 has been translated into many languages and administered in many nations. Consequently, it is important to ensure the similarity of the factor structure of various SF-36 versions because the factor structure of a measurement represents the theoretical framework of the measurement and has a fundamental influence on scoring, interpretation and further analysis for other purposes. Therefore, in the existing literature, many studies have provided evidence showing that the hypothesized factor structure of the SF-36 (eight first-order factors and two second-order factors) was sustained across various versions. For example, when developing a new language version, developers usually conducted exploratory factor analysis (EFA) to ensure that the factor structure of a new version is similar to the original version [2,3]. In addition, several studies directly analysed data from different nations to examine if the hypothesized structure of the SF-36 was sustained across various versions. For instance, Ware *et al*. [4] used EFA to analyse product-moment correlations among the eight subscales for 10 countries in order to test the generalizability of the two-factor structure.

However, the factor analysis done in most of these studies only focused on the two-factor structure (physical and mental health) among the eight subscales, not on the eight-factor structure among the 35 items. In de Vet *et al*.'s study [5], factor analysis [EFA and/or CFA (confirmatory factor analysis)] of the SF-36 of 28 studies was surveyed, showing that results of only six studies could reach a first-order factor analysis for the 35 items, and results of just 25 studies could reach a second-order factor analysis for the eight subscales, of which three achieved both first-order and second-order factor analyses. Most studies conducted factor analysis for the eight subscales, not for the 35 items. However, a procedure that merely focuses on the second-order factor for the eight subscales has its drawbacks. Specifically, if the first-order factor structure among the 35 items is not consistent with the hypothesized structure, items that did not load on its hypothesized factor cannot tap the meaning it is aimed to measure. Therefore, in this situation, it is unreasonable to compute the scores of the eight subscales according to the hypothesized structure. Accordingly, if the scores of the eight subscales, based on the hypothesized structure, were still submitted to factor analysis, the result of second-order factor structure might be distorted. Therefore, in examining the factor structure of the SF-36, it is better to examine the entire factor structure for items and subscales.

In the current study, we examined the entire factor structure of the SF-36 using the SF-36 Taiwan version. The SF-36 Taiwan version was developed in 1996 and has been widely used in Taiwan [6]. However, the factor structure of the SF-36 Taiwan version has not been carefully investigated. Thus, in the current study, both EFA and CFA were conducted to examine the factor structure of the SF-36 Taiwan version. The reason for performing both EFA and CFA is to find a general factor pattern by EFA and cross-validate the EFA results by CFA. Because we did not know if the factor structure of the SF-36 Taiwan version is exactly the same as the hypothesized structure of the SF-36, we decided to perform EFA first, rather than conduct CFA to examine the hypothesized structure of the SF-36 directly.

In the existing literature, evidence can be found which shows that EFA results have not been consistent with the hypothesized structure. For example, in Thumboo *et al.*'s study [7], they found that their EFA results of the English (UK) version and Chinese (HK) version of the SF-36 for the Singapore sample were not consistent with the hypothesized factor structure of the SF-36, although the EFA results of the English (UK) version and Chinese (HK) version were the same. Specifically, the results showed that there were seven, not eight, factors among the 35 items. According to these results, the SF items were not extracted as a distinct factor. In the seven-factor structure, one SF item was factored with MH items; the other SF items were factored with RE items. In addition, two VT items and three MH items were factored as a single factor, and the other two VT items and the other two MH items formed another factor. Their results revealed that the performances of SF, VT and MH scales were not consistent with the hypothesized factor structure. However, although differences between the observed factor structure and the hypothesized factor structure were found, they were not strong enough to determine that the observed factor structure was better than the hypothesized factor structure in interpreting the relationships among the SF-36 items. If the observed factor structure and the hypothesized factor structure can be compared in a CFA framework, we can draw a stronger conclusion of which structure is better. In this way, the observed factor structure can also be cross-validated in a CFA model. Thus, in order to exhaustively investigate the factor structure of the SF-36 Taiwan version, in the current study, EFA was performed first, and CFA was performed to cross-validate the EFA results.

Specifically, in the EFA section, we examined our data in an exploratory perspective to find the factor structure in our sample. First, we performed EFA at item level to establish the factor structure among the 35 items for the eight subscales. Then, based on that result, subscale scores were computed to investigate the second-order factor structure among the eight subscales. According to the EFA results, we could find the exploratory hierarchical factor structure of the SF-36. Further, a CFA was conducted to cross-validate the results from EFA. We decided that if the results of EFA were consistent with the hypothesized factor structure of the SF-36, then the CFA model, based on the hypothesized factor structure, would be built and examined directly. However, if the results of EFA were not consistent with the hypothesized factor structure of the SF-36, a CFA model based on the EFA result would be compared with the CFA model based on the hypothesized factor structure to see (1) if the EFA results could be supported in a CFA model and (2) which model would be more appropriate for the SF-36 Taiwan version.

In this study, general population data from the 2001 National Health Interview Survey (NHIS) in Taiwan was used to examine the factor structure of the SF-36 Taiwan version. The NHIS was conducted by stratified multistage systematic sampling, resulting in 19 777 valid responses. With this abundance of data, we were able to randomly divide the 19 777 responses into two samples for conducting EFA and then CFA. With two independent analyses on two independent samples, we were able to make a convincing argument on the factor structure of the SF-36 Taiwan version.

Methods

Data description

The data in the present study were taken from results of the 2001 NHIS conducted by the National Health Research Institute and the Bureau of Health Promotion in the Department of Health of Taiwan [8]. The 2001 NHIS was intended to provide nationwide estimates of health conditions, health behaviours and usage of medical resources for the population of Taiwan. Using multistage sampling proportional to the size of household populations, 27 160 eligible people living in 7357 households were identified on 16 January 2001. The sample was representative of the national population in age, gender and in terms of the urbanization index [9]. Then, between late August 2001 and January 2002, interviews to collect data for the 2001 NHIS data were conducted. The response rates of 25 464 interviewees living in 6721 households were 93.8% and 91.4%, respectively [10]. To ensure consistency of the interviews, all interviewers received a one-hour training session which included an item-by-item explanation of the SF-36 Taiwan version in Taiwanese. Thirteen senior staff of the Bureau of Health Promotion closely supervised the interviews, reviewed all completed questionnaires, and verified interviewees' responses through random follow-up telephone calls. In addition, they made a cross-item comparison between the scale and other corresponding items in the NHIS. No response by proxy was allowed, even if an interviewee was frail, mentally ill, or unable to communicate. A total of 19 777 (77.7%) participants have completed responses on the SF-36 Taiwan version from 2001 NHIS dataset. In this sample, 9816 (49.6%) were male and 9960 (50.4%) were female. The mean age was 39.33 years (SD = 17.65; nine individuals did not report their ages). In the following analyses, EFA and CFA were conducted using two independent samples derived randomly from the 19 777 participants. The first sample included 9856 participants and was used to conduct EFA (referred to as *the EFA sample*). The second sample included 9921 participants and was used to conduct CFA (referred to as *the CFA sample*). The demographic data for all participants and the two samples are presented in Table 1. According to

Table 1 Demographic data for all participants and the two samples

Demographic variables	All participants*	EFA sample*	CFA sample
Number of participants (*n*)	19 777	9856	9921
Gender [*n* (%)]			
Male	9816 (49.6)	4846 (49.2)	4970 (50.1)
Female	9960 (50.4)	5009 (50.8)	4951 (49.9)
Age (years) (mean ± SD)	39.33 ± 17.65	39.47 ± 17.57	39.18 ± 17.72
12–20	3305 (16.7)	1671 (17.0)	1634 (16.5)
21–30	3843 (19.4)	1957 (19.9)	1886 (19.0)
31–40	3861 (19.5)	1917 (19.5)	1944 (19.6)
41–50	3713 (18.8)	1819 (18.5)	1894 (19.1)
51–60	2183 (11.0)	1058 (10.7)	1125 (11.3)
61+	2863 (14.5)	1428 (14.5)	1435 (14.5)
Missing	9 (0.0004)	6 (0.0006)	3 (0.0003)
Education [*n* (%)]			
Illiterate & primary	3934 (19.9)	1972 (20.0)	1962 (19.8)
Middle school	3854 (19.5)	1873 (19.0)	1981 (20.0)
High school	6316 (31.9)	3213 (32.6)	3103 (31.3)
College & graduate	4394 (22.2)	2183 (22.1)	2211 (22.3)
Missing	1279 (6.5)	615 (6.2)	664 (6.7)
Marital status [*n* (%)]			
Single	6760 (34.2)	3420 (34.7)	3340 (33.7)
Married/living together	10 825 (54.7)	5357 (54.4)	5468 (55.1)
Divorced/separated	1086 (5.5)	511 (5.2)	575 (5.8)
Widowed	1057 (5.3)	539 (5.5)	518 (5.2)
Missing	49 (0.2)	29 (0.3)	20 (0.2)

*One participant was trans-sex.
CFA, confirmatory factor analysis; EFA, exploratory factor analysis; SD, standard deviation.

Table 1, the demographic data for the EFA and CFA samples are quite similar.

Instrument

The *SF-36 Taiwan version* is a multipurpose, short-form health survey. It was developed for the Medical Outcomes Study [11], and has been translated and validated extensively [12]. The SF-36 contains 35 items for measuring eight subscales, including (1) physical functioning (PF), (2) role limitation due to physical problems (RP), (3) bodily pain (BP), (4) general health (GH), (5) vitality (VT), (6) social functioning (SF), (7) role limitation due to emotional problems (RE), and (8) mental health (MH). There is also a single item about perceived change over the past 12 months (out of the eight subscales). Thus, the SF-36 has a total of 36 items. The SF-36 Taiwan version was developed in 1996 and has been widely used in Taiwan [6]. Lu *et al*. [6] examined data quality, scaling assumptions and reliability of the SF-36 Taiwan version on a sample from the database of the NHIS. In terms of scale assumptions, item-scale correlation coefficients ranged from 0.40 to 0.83. Except for MH, the rest of the scales passed item discrimination tests. Finally, the internal consistency reliability reached an acceptable level for all scales ($\alpha > 0.76$), except for SF ($\alpha = 0.65$). Tseng *et al*. [13] also used the NHIS database to establish the norm of the SF-36 Taiwan version. Therefore, the SF-36 Taiwan version has adequate basic psychometric properties for Taiwan people.

Results

Descriptive statistics of items in the SF-36 Taiwan version

Tables 2 and 3 present the descriptive statistics of each item in the SF-36 Taiwan version for the EFA sample and the CFA sample, respectively. The descriptive statistics included mean, standard deviation, skewness and kurtosis. We can readily see that the data of the two samples were not normally distributed. Some items' skewness and kurtosis deviated significantly from normal distribution, especially in the PF subscale. Therefore, in the EFA and CFA procedure that followed, the common method of estimation, based on normal distribution (maximum likelihood), was not adopted. Alternative estimation method which does not rely on a particular distribution assumption was used in the EFA and CFA procedures. In EFA, the iterated principal factor (IPF) extraction method was used to extract factors and estimate factor loadings; in CFA, the asymptotic distribution free (ADF, or the weighted least squares in LISREL software) estimation method was used to estimate parameters in the CFA model. Detailed information is provided in each analysis section for estimating parameters in EFA and CFA models.

Exploratory factor analysis

Exploratory factor analysis was conducted to investigate the factor structure of the SF-36 Taiwan version of the EFA sample. Accord-

Table 2 Item descriptive statistics for EFA sample ($n = 9856$)

Scales	Item	Content	Mean	SD	Skewness	Kurtosis
PF	PF1	Vigorous activities	2.53	0.69	–1.15	–0.02
	PF2	Moderate activities	2.78	0.52	–2.35	4.51
	PF3	Lift/carry	2.87	0.41	–3.19	9.80
	PF4	Climb sev. flights	2.76	0.53	–2.10	3.44
	PF5	Climb one flight	2.92	0.32	–4.38	19.74
	PF6	Bend/kneel	2.83	0.47	–2.73	6.69
	PF7	Walk a mile	2.85	0.44	–3.04	8.59
	PF8	Walk sev. blocks	2.89	0.38	–3.70	13.41
	PF9	Walk one block	2.95	0.27	–5.59	32.84
	PF10	Bathe/dress	2.98	0.18	–8.60	79.12
	Subscale		91.74	16.82	–2.90	9.00
RP	RP1	Cut down time	1.83	0.37	–1.80	1.24
	RP2	Accomplished less	1.83	0.37	–1.79	1.20
	RP3	Limited in kind	1.84	0.37	–1.81	1.26
	RP4	Had difficulty	1.83	0.38	–1.73	0.98
	Subscale		83.23	33.78	–1.78	1.48
BP	BP1	Pain-magnitude	5.11	1.13	–1.10	0.31
	BP2	Pain-interfere	5.08	1.08	–1.08	0.74
	Subscale		81.90	21.16	–0.98	0.29
GH	GH1	EVGFP rating	3.20	1.18	–0.22	–1.28
	GH2	Sick easier	4.15	1.11	–1.28	0.71
	GH3	As healthy	4.05	1.06	–1.07	0.40
	GH4	Health get worse	3.69	1.27	–0.55	–0.90
	GH5	Health excellent	3.92	1.11	–0.96	0.06
	Subscale		70.07	21.85	–0.77	0.08
VT	VT1	Pep/life	4.35	1.28	–0.51	–0.66
	VT2	Energy	4.23	1.29	–0.39	–0.76
	VT3	Worn out	4.65	1.05	–0.81	0.93
	VT4	Tired	4.41	1.12	–0.69	0.67
	Subscale		68.13	18.64	–0.56	0.19
SF	SF1	Social-extent	4.68	0.64	–2.48	7.23
	SF2	Social-time	4.22	0.93	–1.19	1.10
	Subscale		86.29	17.20	–1.62	3.10
RE	RE1	Cut down time	1.80	0.40	–1.46	0.14
	RE2	Accomplished less	1.80	0.40	–1.50	0.24
	RE3	Not careful	1.79	0.41	–1.44	0.09
	Subscale		79.58	36.13	–1.46	0.44
MH	MH1	Nervous	4.44	1.24	–0.80	0.31
	MH2	Down in dumps	4.86	1.02	–1.11	1.69
	MH3	Peaceful	4.45	1.18	–0.58	–0.31
	MH4	Blue/sad	4.82	1.02	–1.00	1.44
	MH5	Happy	4.63	1.21	–0.70	–0.25
	Subscale		72.78	16.61	–0.72	0.68

BP, bodily pain; EFA, exploratory factor analysis; GH, general health; MH, general mental health; PF, physical functioning; RE, role-emotional; RP, role-physical; SD, standard deviation; SF, social functioning; VT, vitality.

ing to the conceptual structure of the SF-36, eight factors were set prior to analysis. Since the data of the EFA sample did not fit normal distribution, the IPF method was chosen to extract factors because it does not require multivariate normality [14]. In addition, because the eight factors of the SF-36 were correlated, the Promax oblique rotation was selected for factor rotation. Table 4 presents the results of EFA with IPF extraction and Promax oblique rotation for the original eight-factor model. Substantial loading was determined by a general rule of 0.30 and is highlighted in Table 4. Table 5 shows the factor correlation matrix among the eight factors.

According to Table 4, it is obvious that the eight factors did not correspond completely with the hypothesized factor structure. Generally, Factor 1 to Factor 7 could be regarded as the PF, MH, RP, GH, RE, VT and BP factors. However, the SF factor was not extracted. The final factor, Factor 8, was also a factor of PF, because substantial loadings of Factor 8 belonged to three PF items. In addition, several items did not fit the structure. For example, in the PF scale, PF1 did not have a substantial loading on Factor 1 as other items, but it had a substantial loading on Factor 8. PF2 and PF4 had substantial loadings on both Factor 1 and Factor 8. In the VT scale, VT3 and VT4 were factored by

Table 3 Item descriptive statistics for CFA sample ($n = 9921$)

Scales	Item	Mean	SD	Skewness	Kurtosis
PF	PF1	2.54	0.68	–1.15	0.02
	PF2	2.79	0.51	–2.37	4.64
	PF3	2.88	0.39	–3.45	11.62
	PF4	2.76	0.53	–2.10	3.43
	PF5	2.93	0.32	–4.60	21.62
	PF6	2.83	0.47	–2.81	7.11
	PF7	2.85	0.44	–3.07	8.77
	PF8	2.90	0.37	–3.81	14.37
	PF9	2.95	0.26	–5.77	34.93
	PF10	2.98	0.18	–8.41	76.60
	Subscale	91.95	16.59	–3.03	10.07
RP	RP1	1.83	0.37	–1.77	1.15
	RP2	1.83	0.38	–1.76	1.09
	RP3	1.84	0.37	–1.85	1.43
	RP4	1.83	0.38	–1.75	1.05
	Subscale	83.25	33.53	–1.78	1.53
BP	BP1	5.11	1.12	–1.12	0.39
	BP2	5.07	1.08	–1.08	0.83
	Subscale	81.77	21.07	–1.00	0.45
GH	GH1	3.19	1.17	–0.20	–1.29
	GH2	4.12	1.12	–1.23	0.56
	GH3	4.04	1.06	–1.06	0.38
	GH4	3.66	1.27	–0.51	–0.96
	GH5	3.91	1.10	–0.98	0.15
	Subscale	69.62	21.90	–0.75	0.07
VT	VT1	4.32	1.29	–0.49	–0.68
	VT2	4.22	1.30	–0.39	–0.76
	VT3	4.65	1.05	–0.76	0.76
	VT4	4.40	1.12	–0.65	0.51
	Subscale	67.91	18.62	–0.55	0.10
SF	SF1	4.67	0.64	–2.41	6.92
	SF2	4.24	0.92	–1.17	1.07
	Subscale	86.36	17.07	–1.59	2.99
RE	RE1	1.79	0.41	–1.45	0.09
	RE2	1.79	0.41	–1.43	0.04
	RE3	1.79	0.41	–1.44	0.06
	Subscale	79.17	36.20	–1.42	0.35
MH	MH1	4.44	1.25	–0.81	0.31
	MH2	4.86	1.02	–1.06	1.48
	MH3	4.45	1.17	–0.55	–0.35
	MH4	4.83	1.00	–0.99	1.46
	MH5	4.62	1.20	–0.70	–0.24
	Subscale	72.76	16.34	–0.65	0.54

BP, bodily pain; CFA, confirmatory factor analysis; GH, general health; MH, general mental health; PF, physical functioning; RE, role-emotional; RP, role-physical; SD, standard deviation; SF, social functioning; VT, vitality.

Factor 2 with MH items, not with VT1 and VT2 on Factor 6. In the SF scale, SF1 did not have any substantial loading on factors, but it had the highest loading on Factor 5 with RE items; SF2 was factorized on Factor 2 with MH items. Finally, in the MH scale, MH3 and MH5 had cross loadings on Factor 2 and Factor 6. These results reveal that the eight-factor structure cannot clearly explain the correlation structure among the items of SF-36 Taiwan version.

Therefore, in order to find a clear factor structure of the SF-36 Taiwan version, a standard EFA procedure was conducted to find appropriate factor numbers and its factor structure. The number of factors was decided by parallel analysis and the interpretability of the factor structure. Previous studies on methods for deciding the number of factors revealed that parallel analysis performs better than other methods, such as eigenvalues above 1 and eye-judged scree plot [14,15]. In parallel analysis, the number of factors is decided by determining how many eigenvalues from the observed correlation matrix (the items of SF-36 Taiwan version) are larger than those from the random correlation matrix. The eigenvalues of the correlation matrix from the SF-36 Taiwan version and the eigenvalues of the random correlation matrix from parallel analysis are plotted in Fig. 1. According to Fig. 1, parallel analysis suggests six factors. However, it is better to analyse different numbers of factors to see which factor structure is more interpretable [16]. Thus, we compared the interpretability of the factor structure on 5, 6 and 7 factors. The seven-factor structure was the best model for interpretation and was closest to the framework of the SF-36. Therefore, the seven-factor structure was finally selected to explain the relationships among the items in the SF-36 Taiwan version.

Table 6 presents the results of EFA with IPF extraction and Promax oblique rotation for the seven-factor model. Substantial loading was determined by a general rule of 0.30 and is highlighted in the table as well. Table 7 shows the factor correlation matrix among the seven factors. In the seven-factor structure, Factor 1 to Factor 7 could be regarded as the PF, MH, RP, GH, RE, BP and VT factors. Compared with the original eight-domain structure of the SF-36, it could be concluded that half of the items of the VT scale (VT3 & VT4) merged into the MH factor and the two items of the SF scale were not extracted as distinct factors; one item of SF2 merged into the MH factor, the other one (SF1) had no substantial loadings. Finally, two items of the MH scale (MH3 & MH5) confounded with two items of the VT scale. This result was exactly the same as the finding in Thumboo *et al.*'s study [7] in Singapore.

Further, it has been proposed that there were hierarchical factor structures underlying the SF-36. That is, beyond the eight first-order factors, there were two second-order factors (physical health and mental health) underlying the eight first-order factors. However, because the first-order factor structure was revised as seven factors in the current analysis, it is worth examining the hierarchical factor structure in this revised model. Therefore, EFA with IPF extraction and Promax oblique rotation was conducted to analyse the correlation matrix among the seven factors in Table 7. The result revealed that there was only one factor underlying the seven factors. Factor loadings of the seven factors ranged from 0.53 to 0.79.

In summary, the EFA results revealed that the original eight-factor structure was not supported. The revised seven-factor structure was better than the original eight-factor structure in describing the correlations among the items of the SF-36 Taiwan version. In addition, hierarchical factor analysis indicated that there was only one second-order factor (general health) underlying the seven first-order factors. This finding was not consistent with the previous findings showing two second-order factors in the SF-36.

Table 4 EFA result of eight-factor model

Scales	Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
PF	PF1	0.26	–0.04	0.05	0.12	0.03	0.03	0.04	0.55
	PF2	0.52	0.01	0.04	0.00	0.02	0.01	–0.03	0.50
	PF3	0.62	0.03	0.05	–0.01	0.01	–0.01	–0.03	0.29
	PF4	0.55	0.02	–0.02	0.03	0.00	0.00	0.01	0.44
	PF5	0.85	–0.02	–0.03	–0.01	0.00	0.01	0.00	0.02
	PF6	0.60	0.00	0.01	–0.01	–0.01	0.02	0.06	0.27
	PF7	0.75	0.03	0.04	–0.02	–0.02	0.00	0.00	0.15
	PF8	0.87	0.01	0.02	–0.01	–0.03	0.00	–0.03	0.04
	PF9	0.99	–0.06	0.01	0.02	–0.03	0.00	–0.01	–0.28
	PF10	0.69	–0.02	–0.02	0.03	0.02	–0.03	0.00	–0.23
RP	RP1	0.01	0.00	0.88	0.01	0.01	0.01	0.00	–0.05
	RP2	0.00	–0.01	0.90	0.01	0.03	0.01	–0.02	–0.05
	RP3	0.05	0.00	0.83	0.00	0.00	0.00	–0.02	0.06
	RP4	0.05	0.01	0.76	0.00	0.01	–0.01	0.03	0.08
BP	BP1	–0.01	0.00	–0.01	0.05	–0.03	0.01	0.83	0.03
	BP2	0.03	0.01	0.02	–0.01	0.01	–0.01	0.95	–0.01
GH	GH1	–0.01	–0.03	0.05	0.41	–0.01	0.16	0.14	0.10
	GH2	0.03	0.09	0.03	0.65	0.00	–0.07	0.02	–0.02
	GH3	0.06	0.03	0.01	0.69	–0.01	–0.02	–0.06	–0.03
	GH4	–0.03	0.04	–0.03	0.57	0.02	0.03	0.01	0.11
	GH5	–0.02	0.02	0.00	0.80	0.00	0.04	–0.01	0.02
VT	VT1	0.01	–0.01	0.03	0.06	0.02	0.73	0.03	0.05
	VT2	–0.01	0.05	0.03	0.01	–0.02	0.81	0.01	0.05
	VT3	–0.01	0.70	0.04	0.04	–0.06	–0.05	0.03	0.05
	VT4	–0.05	0.65	0.04	0.05	–0.05	0.01	0.06	0.02
SF	SF1	0.22	0.15	0.06	0.04	0.26	0.01	0.20	–0.11
	SF2	0.12	0.30	0.08	0.06	0.17	0.04	0.12	–0.06
RE	RE1	–0.02	0.00	–0.01	0.01	0.90	0.00	–0.03	0.03
	RE2	–0.02	–0.02	0.00	0.00	0.92	0.00	–0.03	0.04
	RE3	–0.01	0.06	0.07	–0.02	0.69	–0.01	0.01	–0.01
MH	MH1	–0.06	0.52	–0.02	0.03	0.04	–0.08	0.01	0.02
	MH2	0.04	0.72	–0.03	–0.03	0.03	0.02	–0.05	0.00
	MH3	0.00	0.40	–0.03	–0.02	0.02	0.37	–0.03	–0.10
	MH4	0.02	0.80	–0.01	–0.02	0.02	0.01	–0.04	–0.01
	MH5	0.02	0.36	–0.04	0.06	0.00	0.50	–0.06	–0.07

BP, bodily pain; EFA, exploratory factor analysis; GH, general health; MH, general mental health; PF, physical functioning; RE, role-emotional; RP, role-physical; SF, social functioning; VT, vitality.

Confirmatory factor analysis

Next, a series of CFA was conducted on the CFA sample to cross-validate the results of EFA and to examine the performance of the original model with eight first-order factors and two second-order factors versus the performance of the revised model with seven first-order factors and one second-order factor. However, according to Table 3, the data of the CFA sample did not fit normal distribution; thus, the ADF (the weighted least square in the LISREL) method was used to estimate the parameters in each model. This estimation method was not based on a particular distribution assumption [17] and because the sample size was large enough ($n = 9921$), this method was appropriate for the current sample. LISREL 8.0 [17] was used to analyse the data.

Two models were examined. The first model (Model I) was the original model, which proposed that SF-36 contains eight first-order factors (corresponding to the eight subscales) and two correlated second-order factors. One second-order factor is termed physical health, which influences PF, RP, BP and GH first-order factors. The other one is termed mental health, which influences VT, SF, RE and MH first-order factors. Table 8 presents the standardized estimates of this model, in which all estimates were significant. The chi-square value was 5867.99 (d.f. = 551, $P < 0.01$), rejecting this model. However, because the chi-squared test tends to be influenced by the sample size, a larger sample size (more than 200) may result in significant results (indicating lack of fit); thus, other fit indices needed to be used to evaluate the performance of this model. Therefore, in this study, two incremental fit indices, the non-normed fit index (NNFI) and the comparative fit index (CFI), were chosen to analyse the large sample. The general cut-offs for accepting a model for these two indices were equal to or greater than 0.95 [18]. In addition, two absolute fit indices, the standardized root mean squared residual (SRMR) and the root mean squared error of approximation (RMSEA),

Table 5 Factor correlation matrix in eight-factor model

Factor	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Factor 2	0.28	–					
Factor 3	0.49	0.34	–				
Factor 4	0.42	0.55	0.47	–			
Factor 5	0.31	0.45	0.51	0.36	–		
Factor 6	0.35	0.57	0.35	0.58	0.33	–	
Factor 7	0.46	0.42	0.53	0.56	0.38	0.41	–
Factor 8	0.37	0.18	0.44	0.38	0.16	0.29	0.38

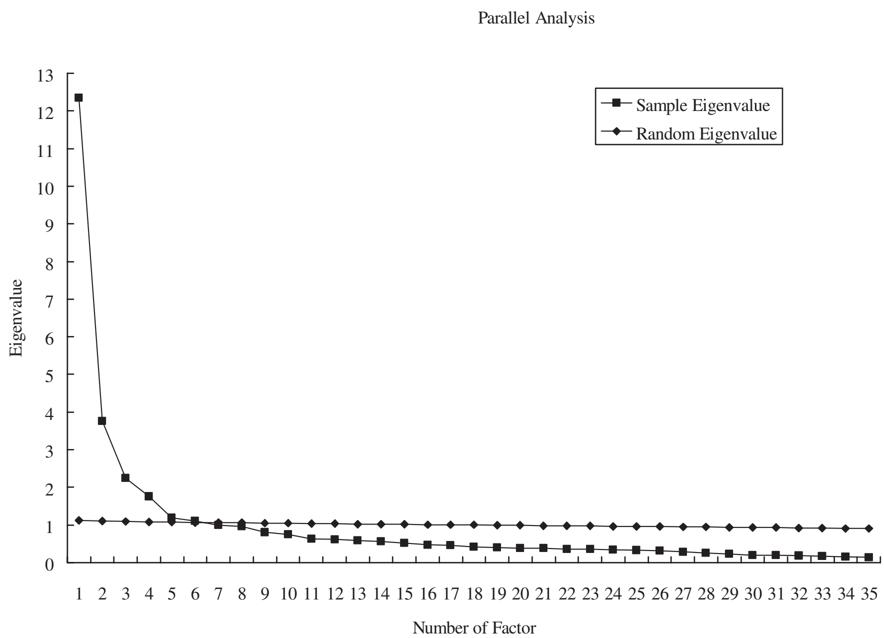

Figure 1 Eigenvalues and parallel analysis in SF-36 scales.

were also used. The general cut-off for accepting a model for RMSEA was equal to or less than 0.05, and the cut-off for SRMR was less than 0.08 [18]. Values of these fit indices for the original model were: NNFI = 0.99, CFI = 0.99, RMSEA = 0.031, SRMR = 0.23. This result revealed that, except for the SRMR, the other three indices indicated a good fit for the original model. However, the correlation between the two second-order factors was 0.93, implying that differentiating the two factors may not be necessary.

The second model (Model II) was the revised model with seven first-order factors and one second-order factor that derived from the EFA result in Table 6. Table 9 presents the standardized estimates of this model, in which all estimates were significant. The chi-square value of this model was 6054.15 (d.f. = 551, $P < 0.01$). In addition, for the revised model, NNFI = 0.99, CFI = 0.99, RMSEA = 0.032 and SRMR = 0.24. This result also revealed that, except for the SRMR, the other three indices indicated a good fit for the revised model.

Although the values of fit indices for the two models were quite similar and indicated an acceptable fit for each, in order to compare their performance, the fit indices for model comparison, the Akaike Information Criterion (AIC) and the Consistent Akaike Information Criterion (CAIC), were used to examine which model was better [17]. The model with lower AIC and CAIC values was found to be better than the other. According to the results, the values of AIC and CAIC for the original model (Model I) were 6025.99 and 6673.98, respectively; and the values of AIC and CAIC for the revised model (Model II) were 6212.15 and 6860.14, respectively. Both the indices indicated that the original model was better than the revised model. The values of various fit indices for the two models are summarized in Table 10.

Discussion

In the current study, factor structure of the SF-36 was investigated by EFA and CFA. As mentioned earlier, many factor analysis studies of the SF-36 rarely examined the whole structure of the SF-36, neglecting items for first-order factor and subscales for second-order factor, and also rarely conducted both EFA and CFA to examine the factor structure of the SF-36. Thus, the strength of our study lies in its use of statistical methods to make a strong conclusion on factor structure of the SF-36.

Next, we shall discuss two main areas for further study, including (1) the characteristics of the empirical data of the SF-36 Taiwan version and its influence on applying EFA and CFA; and (2) the factor structure of the SF-36 Taiwan version based on the EFA and CFA results in the current study and its influence on the psychometric and empirical studies of the SF-36.

Table 6 EFA result of seven-factor model

Scales	Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
PF	PF1	0.36	–0.17	0.20	0.26	–0.04	0.04	0.13
	PF2	0.62	–0.11	0.17	0.13	–0.04	–0.03	0.10
	PF3	0.70	–0.03	0.11	0.04	–0.02	–0.04	0.03
	PF4	0.65	–0.08	0.09	0.14	–0.05	0.01	0.08
	PF5	0.91	0.02	–0.08	–0.06	0.02	0.00	–0.02
	PF6	0.68	–0.05	0.05	0.04	–0.04	0.06	0.06
	PF7	0.83	0.02	0.03	–0.02	–0.03	–0.01	0.01
	PF8	0.94	0.04	–0.03	–0.05	–0.01	–0.04	–0.03
	PF9	0.96	0.07	–0.13	–0.11	0.04	–0.01	–0.09
	PF10	0.68	0.08	–0.13	–0.07	0.07	0.00	–0.11
RP	RP1	–0.04	0.05	0.89	–0.05	0.03	–0.01	–0.03
	RP2	–0.05	0.04	0.91	–0.05	0.05	–0.02	–0.02
	RP3	0.02	0.02	0.90	–0.03	0.00	–0.03	0.00
	RP4	0.03	0.01	0.83	–0.02	0.00	0.02	–0.01
BP	BP1	–0.03	–0.02	–0.02	0.04	–0.05	0.87	0.02
	BP2	0.01	0.00	0.00	–0.04	0.00	0.99	–0.01
GH	GH1	–0.02	–0.04	0.03	0.46	–0.01	0.13	0.16
	GH2	0.00	0.12	–0.02	0.70	0.02	0.01	–0.12
	GH3	0.03	0.08	–0.06	0.72	0.02	–0.07	–0.07
	GH4	–0.04	0.02	–0.05	0.66	0.02	0.00	0.02
	GH5	–0.05	0.05	–0.06	0.86	0.02	–0.02	0.00
VT	VT1	–0.01	0.01	–0.01	0.02	0.04	0.03	0.78
	VT2	–0.04	0.07	0.00	–0.04	–0.01	0.01	0.87
	VT3	0.00	0.67	0.09	0.08	–0.10	0.03	–0.05
	VT4	–0.04	0.63	0.08	0.08	–0.08	0.06	0.00
SF	SF1	0.21	0.18	0.01	–0.01	0.28	0.21	–0.03
	SF2	0.12	0.33	0.06	0.04	0.17	0.12	0.01
RE	RE1	–0.01	–0.05	0.00	0.03	0.89	–0.04	0.02
	RE2	–0.01	–0.07	0.01	0.03	0.92	–0.03	0.02
	RE3	–0.01	0.04	0.08	–0.01	0.69	0.00	0.00
MH	MH1	–0.05	0.50	0.02	0.07	0.01	0.01	–0.09
	MH2	0.05	0.71	0.00	–0.01	0.00	–0.05	0.01
	MH3	–0.02	0.44	–0.07	–0.07	0.03	–0.03	0.36
	MH4	0.03	0.79	0.02	0.00	–0.01	–0.04	–0.01
	MH5	0.00	0.40	–0.09	0.01	0.01	–0.07	0.49

BP, bodily pain; EFA, exploratory factor analysis; GH, general health; MH, general mental health; PF, physical functioning; RE, role-emotional; RP, role-physical; SF, social functioning; VT, vitality.

Table 7 Factor correlation matrix in seven-factor model

Factor	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Factor 2	0.27	–					
Factor 3	0.59	0.32	–				
Factor 4	0.53	0.52	0.58	–			
Factor 5	0.32	0.48	0.50	0.37	–		
Factor 6	0.53	0.43	0.59	0.63	0.41	–	
Factor 7	0.45	0.57	0.45	0.68	0.34	0.49	–

First, regarding the characteristics of the empirical data of the SF-36 Taiwan version, it is obvious that the data of the SF-36 Taiwan version did not fit normal distribution. According to the current analysis, several items had serious problems because of high negative skewness and large positive kurtosis, especially items in the PF subscale. This distribution characteristic may suggest that using these items to capture individual difference for people who have scores in a particular range is not suitable. For example, a highly negative skewed item may be more difficult to use to detect the differences among individuals with a higher score on that item, because a highly negative skewed distribution usually indicates those who have a higher score above the mean on that item. In addition, an item with large positive kurtosis is less able to be used to differentiate individuals in a particular range of scores on a

Table 8 Standardized estimates of factor loadings for the original model

First-order factor loadings								
Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
PF1	0.95	–	–	–	–	–	–	–
PF2	0.99	–	–	–	–	–	–	–
PF3	0.98	–	–	–	–	–	–	–
PF4	0.96	–	–	–	–	–	–	–
PF5	0.99	–	–	–	–	–	–	–
PF6	0.96	–	–	–	–	–	–	–
PF7	0.98	–	–	–	–	–	–	–
PF8	1.00	–	–	–	–	–	–	–
PF9	0.99	–	–	–	–	–	–	–
PF10	0.94	–	–	–	–	–	–	–
RP1	–	0.99	–	–	–	–	–	–
RP2	–	0.99	–	–	–	–	–	–
RP3	–	0.99	–	–	–	–	–	–
RP4	–	0.99	–	–	–	–	–	–
BP1	–	–	0.91	–	–	–	–	–
BP2	–	–	0.99	–	–	–	–	–
GH1	–	–	–	0.82	–	–	–	–
GH2	–	–	–	0.86	–	–	–	–
GH3	–	–	–	0.82	–	–	–	–
GH4	–	–	–	0.77	–	–	–	–
GH5	–	–	–	0.93	–	–	–	–
VT1	–	–	–	–	0.88	–	–	–
VT2	–	–	–	–	0.90	–	–	–
VT3	–	–	–	–	0.85	–	–	–
VT4	–	–	–	–	0.84	–	–	–
SF1	–	–	–	–	–	0.90	–	–
SF2	–	–	–	–	–	0.85	–	–
RE1	–	–	–	–	–	–	0.97	–
RE2	–	–	–	–	–	–	1.00	–
RE3	–	–	–	–	–	–	0.92	–
MH1	–	–	–	–	–	–	–	0.53
MH2	–	–	–	–	–	–	–	0.81
MH3	–	–	–	–	–	–	–	0.73
MH4	–	–	–	–	–	–	–	0.87
MH5	–	–	–	–	–	–	–	0.86

Second-order factor loadings and factor correlation		
	Physical health	Mental health
PF	0.90	–
RP	0.99	–
BP	0.88	–
GH	0.93	–
VT	–	0.97
SF	–	0.96
RE	–	0.86
MH	–	0.92
Corr	0.93	

BP, bodily pain; Corr, factor correlation between two second-order factors; GH, general health; MH, general mental health; PF, physical functioning; RE, role-emotional; RP, role-physical; SF, social functioning; VT, vitality.

particular item, because positive kurtosis indicates a relatively peaked distribution and people with scores in the range of the peak area are not easily differentiated according to their score. Of course, the observed distribution in the data was not the characteristics of the SF-36 Taiwan version itself, but was the characteristics of the sample that completed the SF-36 Taiwan version. A different sample may yield a different distribution. We think that the characteristic of non-normal distribution for the current general population lays in the content of the SF-36. For example, in the SF-36, items on physical functioning tapped the ability to perform basic physical activities. For the general population, most people can easily accomplish these activities and endorse a high score on these items,

Table 9 Standardized estimates of factor loadings for the revised model

First-order factor loadings							
Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
PF1	0.94	–	–	–	–	–	–
PF2	0.99	–	–	–	–	–	–
PF3	0.98	–	–	–	–	–	–
PF4	0.96	–	–	–	–	–	–
PF5	1.00	–	–	–	–	–	–
PF6	0.95	–	–	–	–	–	–
PF7	0.98	–	–	–	–	–	–
PF8	0.99	–	–	–	–	–	–
PF9	0.99	–	–	–	–	–	–
PF10	0.94	–	–	–	–	–	–
RP1	–	0.99	–	–	–	–	–
RP2	–	1.00	–	–	–	–	–
RP3	–	0.99	–	–	–	–	–
RP4	–	0.99	–	–	–	–	–
BP1	–	–	0.89	–	–	–	–
BP2	–	–	0.97	–	–	–	–
GH1	–	–	–	0.81	–	–	–
GH2	–	–	–	0.87	–	–	–
GH3	–	–	–	0.81	–	–	–
GH4	–	–	–	0.76	–	–	–
GH5	–	–	–	0.92	–	–	–
VT1	–	–	–	–	0.90	–	–
VT2	–	–	–	–	0.93	–	–
VT3	–	–	–	–	–	0.85	–
VT4	–	–	–	–	–	0.84	–
SF1	–	–	–	–	–	–	0.90
SF2	–	–	–	–	–	0.85	–
RE1	–	–	–	–	–	–	0.98
RE2	–	–	–	–	–	–	1.00
RE3	–	–	–	–	–	–	0.92
MH1	–	–	–	–	–	0.54	–
MH2	–	–	–	–	–	0.80	–
MH3	–	–	–	–	0.43	0.31	–
MH4	–	–	–	–	–	0.86	–
MH5	–	–	–	–	0.56	0.31	–
Second-order factor loadings							
	General health						
Factor1	0.89						
Factor2	0.98						
Factor3	0.88						
Factor4	0.92						
Factor5	0.86						
Factor6	0.91						
Factor7	0.94						

BP, bodily pain; GH, general health; MH, general mental health; PF, physical functioning; RE, role-emotional; RP, role-physical; SF, social functioning; VT, vitality.

and then, a highly negative skewed distribution of item scores results. Thus, we cannot blame the SF-36 Taiwan version according to the non-normal distribution in a particular sample or population. However, it is fair to say that several items of the SF-36 Taiwan version were not appropriate for the current (general) population, because these items did not have ability to capture individual difference fairly in the current population.

Further, the non-normal distribution of the data may also give rise to inadequate statistical analysis. Generally, several commonly used methods or procedures, such as ANOVA, linear regression, EFA and CFA, and so on, usually are based on normal distribution theory. If data distributions seriously deviate from normal distribution, the statistical results based on normal distribution theory cannot be sustained. Therefore, in the current study,

Table 10 Values of fit indices for two models ($n = 9921$)

Model	d.f.	χ^2	CFI	NNFI	RMSEA	SRMR	AIC	CAIC
Model I	551	5867.99	0.99	0.99	0.031	0.23	6025.99	6673.98
Model II	551	6054.15	0.99	0.99	0.032	0.24	6212.15	6860.14

Note. Model I contains 35 items, in which eight first-order factors and two second-order factors were specified. Model II contains 35 items, in which seven first-order factors and one second-order factor were specified.

AIC, the Akaike Information Criterion; CAIC, the Consistent Akaike Information Criterion; CFI, comparative fit index; d.f., degree of freedom; NNFI, non-normed fit index; RMSEA, root mean squared error of approximation; SRMR, standardized root mean squared residual.

we cannot adopt the maximum likelihood method (which has normal distribution assumption) to perform EFA and CFA. Thus, we chose the IPF extraction method in EFA, and the ADF estimation method in CFA to analyse the non-normal data. However, it should be noted that in EFA, the IPF extraction method is fine for a common sample size, but to apply the ADF estimation method in CFA, a very large sample size (as the size of the current study) is needed [17]. Accordingly, we recommend researchers to confirm data distribution before conducting further analyses of the SF-36. This verification procedure would help researchers to choose appropriate methods.

Regarding the second issue concerning the factor structure of the SF-36 Taiwan version, our EFA results revealed that there were only seven first-order factors and one second-order factor underlying the SF-36 Taiwan version. The seven-first-order-factor structure was consistent with the finding in Thumboo *et al*.'s study in Singapore [7]. They examined the factor structure of the SF-36 English (UK) version and Chinese (HK) version by EFA. At the item level, they showed that half of the items of the VT scale (VT3 & VT4) merged into the MH factor, and the two items of the SF scale were not extracted as a distinct factor. One item (SF2) emerged as the MH factor; the other one (SF1) had no substantial loadings, but was regarded on the RE factor because its largest loading was on the RE factor. Finally, two items for the MH scale (MH3 & MH5) confounded with two items of the VT scale. The result was exactly the same as the EFA result in the current study and may suggest that this seven-factor structure is reliable across different samples and different versions. However, Thumboo *et al.* [7] used the eight subscales scores (based on the original eight-factor structure) to examine the second-order factor structure; they did not use the correlations among the seven first-order factors to examine the second-order factor structure. As mentioned earlier, if the first-order factor structure is not consistent with the original eight-factor structure, it is unreasonable to use the eight subscales scores to examine the second-order factor structure. Therefore, in our study, we examined the second-order factor structure based on the seven-first-order-factor structure, and the result showed that there was only one second-order factor underlying the SF-36 Taiwan version.

However, the CFA results revealed that the original model and the revised model had similar performances on the fit indices of NNFI, CFI, RMSEA and SRMR. In addition, AIC and CAIC indices indicated that the original model with eight first-order factors and two second-order factors was better than the revised model with seven first-order factors and one second-order factor. This finding was not consistent with the finding from the EFA, in which the eight-factor structure in the original model was not supported. This inconsistency may have resulted from the theoretical differences between EFA and CFA methods. We know that EFA aims to explore a reasonable factor structure from a correlation matrix and CFA aims to examine a specific factor structure from a covariance matrix. Therefore, it is possible that EFA and CFA would not have the same results, as was the case in this study. Nevertheless, the conclusion drawn from the CFA approach was more desirable, because it provided a statistical basis to compare different models. Thus, taking all of the results into account, we can only conclude that the original model was acceptable, although EFA results also showed an alternative model. In addition, the values of SRMR for the two CFA model also revealed that there is a possible alternative model for the SF-36. This is because the SRMR index is sensitive to model misspecification [18]. A large value of the SRMR indicates that there is something wrong with the model specification. Since we did not have prior theory to modify the structure model of the SF-36, we did not re-specify the model to reduce the value of SRMR. However, the information of SRMR provided a hint for further investigation on the factor structure of the SF-36.

Generally, according to the EFA and CFA results, we would like to mention that the factor structure of the SF-36 should be further examined by another sample for cross-validation. As we found in the EFA results, the eight items, including VT3, VT4, SF1, SF2, MH3 and MH5, did not precisely load on their posited factors in the eight-factor structure. Although the original model with eight first-order factors and two second-order factors is still acceptable, more extensive work on the psychometric study of the SF-36 should be conducted to clarify the meaning of these items in future studies.

Acknowledgements

This study was supported by the National Science Council (NSC 94-2413-H-002-018) and the National Health Research Institute (NHRI-EX94-9204PP). We thank the National Health Research Institute and the Bureau of Health Promotion, Department of Health, Taiwan for providing the NHIS data.

References

1. Ware, J. E. Jr & Gandek, B. (1998) Overview of the SF-36 health survey and the IQOLA project. *Journal of Clinical Epidemiology*, 51, 903–912.
2. Lam, C. L., Gandek, B., Ren, X. S. & Chan, M. S. (1998) Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 health survey. *Journal of Clinical Epidemiology*, 51, 1139–1147.

3. Li, L., Wang, H. M. & Shen, Y. (2003) Chinese SF-36 health survey: translation, cultural adaptation, validation, and normalization. *Journal of Epidemiology and Community Health*, 57, 259–263.
4. Ware, J. E. Jr, Kosinski, M., Gandek, B., *et al.* (1998) The factor structure of the SF-36 health survey in 10 countries: results from the IQOLA project. *Journal of Clinical Epidemiology*, 51, 1159–1165.
5. Vet, H. C. W., Adèr, H. J., Terwee, C. B. & Pouwer, F. (2005) Are factor analytical techniques used appropriately in the validation of heath status questionnaire? A systematic review on the quality of factor analysis of the SF-36. *Quality of Life Research*, 14, 1203–1218.
6. Lu, J. R., Tseng, H. M. & Tsai, Y. J. (2003) Assessment of health-related quality of life in Taiwan (I): development and psychometric testing of SF-36 Taiwan version. *Taiwan Journal of Public Health*, 22, 501–511 (in Chinese).
7. Thumboo, J., Fong, K. Y., Machin, D., Chan, S. P., Leon, K. H., Feng, P. H., Thio, S. T. & Boe, M. L. (2001) A community-based study of scaling assumptions and construct validity of the English (UK) and Chinese (HK) SF-36 in Singapore. *Quality of Life Research*, 10, 175–188.
8. Shih, Y. T., Chang, H. Y., Le, K. H., Lin, M. C. & Su, W. C. (2002) Introduction to the National Health Interview Survey (NHIS). *NHIS Brief Communication*, No. 1, June.
9. Hung, Y. T. (2002) Sampling design of the National Health Interview Survey. *NHIS Brief Communication*, No. 2, June.
10. Lin, S. H. (2002) Field collection and completeness of data in the National Health Interview Survey. *NHIS Brief Communication*, No. 1, June.
11. Ware, J. E. & Sherbourne, C. (1992) The MOS 36 item short form health survey. I. Conceptual framework and item selection. *Medical Care*, 30, 473–483.
12. Gandek, B. & Ware, J. E. Jr (1998) Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. *Journal of Clinical Epidemiology*, 51, 953–959.
13. Tseng, H. M., Lu, J. R. & Tsai, Y. J. (2003) Assessment of health-related quality of life (II). Norming and validation of SF-36 Taiwan version. *Taiwan Journal of Public Health*, 22, 512–518 (in Chinese).
14. Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. & Strahan, E. J. (1999) Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272–299.
15. Zwick, W. R. & Velicer, W. F. (1986) Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432–442.
16. Ford, J. K., MacCallum, R. C. & Tait, M. (1986) The application of exploratory factor analysis in applied psychology: a critical review and analysis. *Personnel Psychology*, 39, 291–314.
17. Joreskog, K. G. & Sorbom, D. (1993) LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. Hillsdale, NJ: Lawrence Erlbaum Associates.
18. Hu, L. & Bentler, P. M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.

