Can the Web-Form WHOQOL-BREF be an Alternative to the Paper-Form?

**Wen-Ching Chen · Jung-Der Wang · Jing-Shiang Hwang ·
Chiao-Chicy Chen · Chia-Huei Wu · Grace Yao**

Abstract The purpose of this study was to test whether the web version is an alternative to the paper version of the short version of the World Health Organization Quality of Life assessment (WHOQOL-BREF). Two studies were conducted. Study 1 used crossover self-controlled trials with 80 participants to compare the web and paper versions and to determine the test–retest reliability of the web version. Study 2 used data from 1,016 web participants to analyze the internal consistency and concurrent and construct validity of the web version. The correlations of domain scores between the web and paper versions ranged from 0.71 to 0.85. Dependent t tests showed no significant differences in domain scores between these two versions. The intra-class correlation coefficients (ICC) for test–retest reliability of web version ranged from 0.79 to 0.91. The Cronbach's α for internal consistency reliability ranged from 0.60 to 0.83. Multiple regression models indicated that the web version has good concurrent validity. Confirmatory factor analysis (CFA) for the second-order hierarchical factor model also supported the construct validity of the web

W.-C. Chen · J.-D. Wang
Institute of Occupational Medicine and Industrial Hygiene, College of Public Health,
National Taiwan University, Taipei, Taiwan

W.-C. Chen
Yuli Hospital, Department of Health, Executive Yuan, Taiwan

J.-D. Wang
Department of Internal Medicine and Department of Environmental and Occupational Medicine,
National Taiwan University Hospital, Taipei, Taiwan

J.-S. Hwang
Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

C.-C. Chen
Song De Branch, Taipei City Hospital, Taipei, Taiwan

C.-H. Wu · G. Yao (✉)
Department of Psychology, College of Science, National Taiwan University,
No. 1, Sect. 4, Roosevelt Road, Taipei 10617, Taiwan
e-mail: kaiping@ntu.edu.tw

version. The web version of the WHOQOL-BREF can be the alternative to the paper version for health-related quality of life (HR-QOL) evaluation.

Keywords Comparison · Reliability · Validity · Web version · WHOQOL-BREF

1 Introduction

The measurement of health-related quality of life (HR-QOL) has become increasingly important in outcome research promoting patient's value in healthcare services (Institute for Strategy and Competitiveness, Harvard Business School 2007; International Society for Quality of Life Research 2007; Porter and Teisberg 2006). As the US Food and Drug Administration (FDA) began considering to include patient-reported outcome as part of clinical trials (US Department of Health and Human Services, Food and Drug Administration 2007), it is likely that many generic and condition specific questionnaires used as the instruments to assess the outcome, will be in progressively greater use in health care research in the future (McKenna and Doward 2004). Many reports have noticed disadvantages of using a paper questionnaire, including data-collecting errors, missing data, time consumption, and manual errors in coding or typing data into the database (Bushnell et al. 2006; Pouwer et al. 1998; Wright et al. 1998). In contrast to the paper version, a web- or computer-based survey seems to preserve more privacy (Gordon et al. 2003; Wang et al. 2005), more effective in collection of unfavorable clinical events (Bush et al. 2005), and more efficient for repeated follow-up of patients (Mullen et al. 2004). Feasibility and reliability of data have also been confirmed (Mullen et al. 2004; Kimura et al. 2005; Lin 2003; Carlbring et al. 2007). The task force of International Society of Pharmacoeconomics and Outcomes Research, after careful consideration of future impacts, has been promoting the electronic modes of administration for patient-reported outcome measures (The International Society for Pharmacoeconomics and Outcomes Research 2007). It is likely that Internet- and computer-based studies will increase in the future, especially studies requiring repeated measurements and long-term follow-up for the full cycle of a medical condition (Institute for Strategy and Competitiveness, Harvard Business School 2007).

The short version of the World Health Organization Quality of Life assessment (WHOQOL-BREF), which was simplified from the WHOQOL-100, is one of the leading generic measurements of HR-QOL because it consists of items concerned with the meaning to respondents of different concepts of life, has good psychometric properties, and is available in most major languages (Skevington et al. 2004). Traditionally, the WHOQOL-BREF has been applied in a pencil and paper format. Although there were several studies using Internet or computer based format for WHOQOL-100 (Mason et al. 2004) or WHOQOL-BREF (Fellinger et al. 2005), the reliability and validity of the data obtained from electronic system were not specifically mentioned.

To test whether the web format of the WHOQOL-BREF is equivalent to the paper format, we conducted following two studies: the first used a cross-over randomized self-controlled trial to check the correlation and mean differences between the web and paper versions and the test–retest reliability and mean differences across time of the web version. The other used data collected through the Internet to test the internal consistency, concurrent validity, and construct validity of the web version.

2 Study 1: Comparing the Web Version of the WHOQOL-BREF with the Paper Version

2.1 Methods

2.1.1 Sample

The study was first approved by the Institutional Review Board of the Song De Branch of Taipei City Hospital before it started. In September 2005, 80 out of 132 nursing staff at this hospital that provides acute treatment for psychotic patients volunteered for the study. Ward-based clusters were randomly divided into two groups, Group 1 and Group 2. In other words, nurses worked at the same ward were randomly assigned to the same group. Frequency distributions of demographics, including gender, age, marital status, and education, physical disease, and health behaviors, including smoking and drinking did not differ significantly between Group 1 and Group 2 according to the Chi-square test (Table 1).

Table 1 Frequency distributions, n (%), of demographics, physical illness, and health behaviors of nurses in Study 1 and the participants in study 2, and the values of the Chi-square test between Groups 1 and 2 in study 1

	Sample in study 1					Sample in study 2
	Group 1	Group 2	DF	Statistic test		
	$n = 38$	$n = 42$		Chi-square value	p-value	$n = 1,016$
Sex			1	3.81	0.051	
Female	38	38				691 (68.0)
Male	0	4				325 (32.0)
Age			1	1.49	0.22	
<40	32	39				671 (66.0)
>40	6	3				345 (34.0)
Marital status			2	5.27	0.07	
Single	20	30				404 (39.8)
Married	17	9				612 (60.2)
Other	1	3				
Education			1	1.14	0.29	
University	34	34				794 (78.2)
High school	4	8				222 (21.8)
Physical illness[a]			1	0.26	0.61	
Yes	9	8				160 (15.7)
No	29	34				856 (84.3)
Smoking			1	0.92	0.34	
Yes	0	1				81 (8.0)
No	38	41				935 (92.0)
Alcohol			1	0.62	0.43	
Yes	14	12				336 (33.1)
No	24	30				690 (66.9)

[a] Physical illness includes hypertension and/or diabetes

2.1.2 Instrument

The WHOQOL group developed the WHOQOL-100 questionnaire in 1995 (The WHO-QOL group 1995, 1998), which is an authentic means to measure a person's feeling of disease impact and impairment on their life. The measure is too long to be used in a clinic, so it was simplified into the WHOQOL-BREF version (Skevington et al. 2004). Now, it is applied in many fields of HR-QOL measurement (Naumann and Byrne 2004; Taylor et al. 2004). The development of the WHOQOL-BREF Taiwan version was completed in 2000 (Yao et al. 2002). It has been broadly used in research (Fang et al. 2002; Hsiung et al. 2005; Lai et al. 2006; Yang et al. 2005). The WHOQOL-BREF is a generic, multilingual profile for subjective measurement of HR-QOL. It contains 26 items. The first and second questions belong to Facet G, which measures global QOL and general health. They are labeled as G1 and G2, respectively, and are examined separately. The other 24 items represent the 24 Facets in the WHOQOL-BREF. These items are classified into four domains: physical, psychological, social, and environmental. The seven items of the physical domain include items 3 (pain), 4 (medication), 10 (energy), 15 (mobility), 16 (sleep), 17 (daily activities), and 18 (work). The six items of the psychological domain are 5 (positive feelings), 6 (spirit), 7 (thinking), 11 (body image), 19 (self-esteem), and 26 (negative feelings). The social domain contains three items, 20 (personal relationships), 21 (sexual satisfaction), and 22 (social support). The remaining eight items belong to the environmental domain: 8 (safety), 9 (physical environment), 12 (finances), 13 (information availability), 14 (leisure), 23 (living environment), 24 (medical services), and 25 (transport).

We followed the standard scoring method for the WHOQOL-BREF (The WHOQOL Group 1993; World Health Organization 1996). Higher scores indicate a better HR-QOL, except items, 3, 4, and 26, which are reversely coded. The score for each domain equals the mean value of its items multiplied by four. For example, the score of physical domain is the mean of the scores on items 3, 4, 10, 15, 16, 17, and 18 multiplied by four. The value of each domain ranges from 4 to 20. With the exception of the social domain, which allows at the most one missing data points, if two (or more) items within a domain are missing, the score of that domain was not calculated. In addition, if more than 20% of the total data was missing, then that participant's data was discarded. If missing data occurred and did not violate these rules, the missing data was replaced with the mean of the other items within that same domain. Past studies have shown that the test–retest reliability coefficients and the internal consistency of the paper version of the WHOQOL-BREF are 0.76–0.86, and 0.70–0.77 at the item and domain levels respectively (Yao et al. 2002).

2.1.3 Procedure

A website, http://www.healthcity.net.tw/vio_ask/, was set up for the WHOQOL-BREF. The contents and sequence of the questions were exactly the same as the paper version. After the 80 volunteers were enrolled and consent forms signed, each was provided an account number and password to access the website. When participants were completing the questionnaire, they would see two or three questions displayed on the screen. If they did not complete all the questions before they sent out the data, the screen would return to the questions left blank and prompt for completion. Once the data was successfully sent out, it was automatically transferred to an Excel® database.

For the counter balance reason, we used the crossover study design. The 38 nurses in Group 1 were asked to complete the web version first and then the paper version and the

other 42 nurses in Group 2 did the paper version first, then the web version 2 or 3 days later. A total of 34 and 38 participants completed both the web and paper versions in Groups 1 and 2, respectively. To evaluate the test–retest reliability of the web version, 75 nurses who had completed the web version of the WHOQOL-BREF initially were invited to do the web version again 10–14 days later. A total of sixty actually completed the second questionnaire within 2 weeks.

2.1.4 Analysis

We used Pearson's correlation coefficients to compare the domain scores of web and paper versions, and used the intra-class correlation coefficient (ICC) to assess the test–retest reliability of the web version of the WHOQOL-BREF. Furthermore, dependent *t* tests were also performed to see if there is no difference in domain scores between web and paper versions and the test–retest scores of the web version. In addition, Bland and Altman (1986) method also was used to compare these two different versions. The software used for data analysis was SAS® 8.2.

2.2 Results

2.2.1 Missing Data

There were no missing data on the web version because the program was designed to force participants to complete all the questionnaires. However, on the paper version there were four missing responses, one each on items 8, 9, 18, and 22 in Group 1, and six missing responses in Group 2, one on item 12, and five on item 21. The most missing data, five (about 0.48%), were on item 21.

2.2.2 Analysis of Correlation Between the Web and Paper Versions

The correlations between the web and paper versions ranged from 0.71 to 0.83 and 0.76 to 0.85 in Groups 1 and 2, respectively. The lowest value in both groups was the social domain. The highest value for Group 1 was the physical domain, and for Group 2 it was the psychological domain (Table 2). For each domain, the difference in correlation between the two groups was not statistically significant according to the *z* test after Fisher's

Table 2 Pearson's coefficients of correlation between Web-paper form, intra-class correlation coefficients (ICC) of test–retest, Cronbach's α for internal consistency, and inter-item correlations of web-form WHOQOL-BREF

Domain	Study 1			Study 2	
	Web-paper Group 1 ($n = 34$)	Paper-web Group 2 ($n = 38$)	Test–retest, ICC ($n = 60$)	Cronbach's α ($n = 1,016$)	Inter-item correlations (mean) ($n = 1,016$)
Physical	0.83	0.77	0.91 (95%CI: 0.85–0.94)	0.75	0.13–0.58 (0.31)
Psychological	0.79	0.85	0.79 (95%CI: 0.64–0.87)	0.82	0.31–0.53 (0.43)
Social	0.71	0.76	0.91 (95%CI: 0.85–0.94)	0.60	0.29–0.40 (0.34)
Environmental	0.75	0.83	0.89 (95%CI: 0.81–0.93)	0.83	0.26–0.43 (0.38)

transformation. The correlation matrices of the four domain scores across two versions were also display in Appendix for supplements.

2.2.3 Test–Retest Reliability of the Web Version

The values of the intra-class correlation coefficient (ICC) were 0.91, 0.79, 0.91, and 0.89 for the physical, psychological, social, and environmental domains, respectively (Table 2). At the domain level, all values were over 0.75, which are considered acceptable. The lowest value was in the psychological domain. The correlation matrices of the four domain scores across two times were also display in Appendix for supplements.

2.2.4 Mean Difference Tests Between Web and Paper Versions and Test–Retest Scores of Web Version

A 2×2 mix design ANOVA was conducted to test if there are mean differences in four domain scores between the four conditions from by two groups with two versions. Effect of group was treated as between-subject effect and effect of versions was treated as within-subject effect. The interaction effect was also included in analysis. For each domain score, two main effects and the interaction effect were all not significant (all $ps > 0.05$), showing that group, version and their interaction did not result in mean differences on the four domain scores. The same finding was observed when dependent t tests in four domain scores between web and paper versions were directly conducted (see Table 3).

In addition, except for psychological domain, dependent t tests for other three domains in test–retest score of web version were not significant, revealing that participants had higher scores in psychological domain at Time 2 than at Time 1 in the test–retest interval.

It was also found that proportions of differences between two versions in a range of mean difference plus and minus two standard deviation of the difference for each domain using Bland and Altman (1986) method for the two groups were all above 0.94, except for the social domain for Group 1, which only has a value of 0.88. Similarly, in test–retest study, all the results showed were all above 0.92, except for the social domain, 0.88. (Table 4).

3 Study 2: Verification of the Reliability and Validity of the Web Version of the WHOQOL-BREF

3.1 Method

3.1.1 Instrument

The same WHOQOL-BREF instrument was used as in Study 1.

3.1.2 Procedure and Sample

The WHOQOL-BREF was uploaded to a website, http://tpeir.health.gov.tw/survey/, which was accessible to the general public without a password. An inform consent page stating the purpose of this study, the use of data, and guarantee for the privacy of participants will be showed first and it requires their agreement if they like to begin the performance of questionnaire. The computer program and configuration were equivalent to Study 1 except

Table 3 The mean, standard deviation (SD), difference, diff., and its dependent t test value (t) of each domain in study 1

Domain	Group 1 ($n = 34$)			Group 2 ($n = 38$)			Group 3 ($n = 60$)		
	Web mean (SD)	Paper mean (SD)	diff. (t)	Paper mean (SD)	Web mean (SD)	diff. (t)	Test mean (SD)	Pretest mean (SD)	diff. (t)
Physical	14.08 (1.78)	13.95 (1.83)	0.13 (1.06)	14.11 (2.02)	13.83 (2.24)	0.27 (1.45)	13.95 (1.88)	13.80 (1.64)	0.16 (1.48)
Psychological	12.57 (2.17)	12.69 (1.95)	−0.12 (1.36)	12.88 (2.57)	12.98 (2.58)	−0.11 (1.43)	12.61 (2.69)	12.21 (2.49)	0.40 (1.48)*
Social	13.53 (2.23)	13.37 (2.36)	0.16 (1.76)	13.58 (2.14)	13.86 (2.09)	−0.28 (1.46)	13.38 (2.43)	13.22 (2.25)	0.16 (1.37)
Environmental	12.64 (1.90)	12.70 (1.73)	−0.05 (1.28)	12.46 (2.30)	12.78 (2.43)	−0.32 (1.40)	12.70 (2.25)	12.53 (2.20)	0.17 (1.41)

* $p < 0.05$

Table 4 Proportion of differences between two versions in a range of mean difference plus and minus two standard deviation of the difference for each domain using Bland and Altman (1986) method in study 1	Domain	Web-paper Group 1 (n = 34)	Paper-web Group 2 (n = 38)	Test–retest (ICC) (n = 60)
	Physical	0.94	0.97	0.93
	Psychological	0.97	0.95	0.92
	Social	0.88	0.95	0.88
	Environmental	0.97	0.95	0.93

feedback information was shown on the screen immediately after the questionnaire was completed and submitted. The information included some words of thanks for participating in the study and the individual's HR-QOL status as compared to the normative data for the population of Taiwan, which was established using 13,010 cases obtained from a National Health Interview Survey conducted in 2001 (Lin et al. 2003). To encourage people to fill in the questionnaire on the website there was a promotional activity, in which any participant who completed the questionnaire was entered in a drawing to get a beautiful T-shirt. To give people an equal treatment, we also designed a computer program to set a limit that one person can only have one chance to complete the questionnaire within 1 month. If a person has more than one data within the survey period, the first data was used for the analysis. This design supposedly excluded the possibility that a person might complete the questionnaire on multiple occasions, hence contributing to the reliability of the questionnaire.

A total of 1,016 people completed the survey online within 6 months (Nov. 1, 2004 to April 30, 2005). About two-thirds of the participants were under 40. More than half were married, female, and had a university education. Most had no major physical illness and did not smoke or drink (Table 1).

3.1.3 Analysis

The data were automatically coded into a database, which was used for the following analyses: first, we calculated the Cronbach's α for internal consistency reliability of the web version of the WHOQOL-BREF. The accepted minimum standard is 0.7, which implies the homogeneity of content and indicates that the score is free from random error. Second, to obtain concurrent validity, we conducted multiple regression models using G1, G2, or G1 + G2 (the sum of G1 and G2) as the criterion variable and the four domain scores as the predictive variables. Because G1 and G2 are two items measuring global quality of life and health, it was hypothesized that the four domain-specific quality of life measures should show their validity to predict global measures. Hence, we used G1 and G2 and the sum score of these two items to examine the concurrent validity of the four domain measures. Higher R-square, percentage of score variance explained, indicates better predictive validity.

Finally, confirmatory factor analysis (CFA) was performed to determine the construct validity of the web version of the WHOQOL-BREF by LISREL 8.0 (Joreskog and Sorbom 1993) with maximum likelihood estimation. The hierarchical model for the standard WHOQOL-BREF version (Skevington et al. 2004; Yao et al. 2002) was examined with 24 items. In this hierarchical model, a common HR-QOL factor influenced four domain factors, which were indicated by its items. In addition, the unique variances of each variable were uncorrelated. To estimate parameters, factor variance of the common HR-QOL factor was set to be one, and one item loading was set to be one within each

first-order factor. To evaluate the model, four fit indices were used in conjunction with the Chi-square test, which is not an ideal test for model fit because it tend to be significant (indicating lack of fit) when sample size is large. Thus the four fit indices which contained two incremental fit indices (NNFI, CFI) and two absolute fit indices (SRMR, RMSEA) were suggested by Hu and Bentler (1999). Higher values in NNFI and CFI (higher than 0.90 or 0.95) and lower values in RMSEA (lower than 0.08 or 0.05) and SRMR (lower than 0.08) indicate a good fit (Bentler 1990; Browne and Cudeck 1993; Hoyle 1995). In addition, in order to see if the hierarchical model is better than a single factor model (all items were influenced by one factor), values of fit indices for model comparison (AIC, CAIC, and ECVI) were compared between these two models. A model with lower values in these three fit indices is better.

3.2 Results

3.2.1 Reliability of Internal Consistency by Cronbach's α

As summarized in Table 2, the Cronbach's αs were 0.75, 0.82, 0.60 and 0.83 for the physical, psychological, social, and environmental domains, respectively. The lowest value was on the social domain. Inter-item correlations within each domain were acceptable with the means higher than 0.30. However, the two reversed items in the physical domain had lower correlations with other items in the physical domain (ranged from 0.10 to 0.20), which resulted in a lower mean inter-item correlation in the physical domain.

3.2.2 Concurrent Validity by Regressing Coefficients (Standardized β)

The best predictor for all three variables G1, G2, and G1 + G2, was the environmental domain. In this study the β values were 0.33, 0.33, and 0.37 respectively. In addition, all the estimates of β were statistically significant. The results showed that the variance in predicting the G1, G2, and G1 + G2 scores could be explained by the scores of the four domains: 43, 37, and 51% (R^2) respectively (see Table 5).

3.2.3 Construct validity of web version by confirmatory factor analysis (CFA)
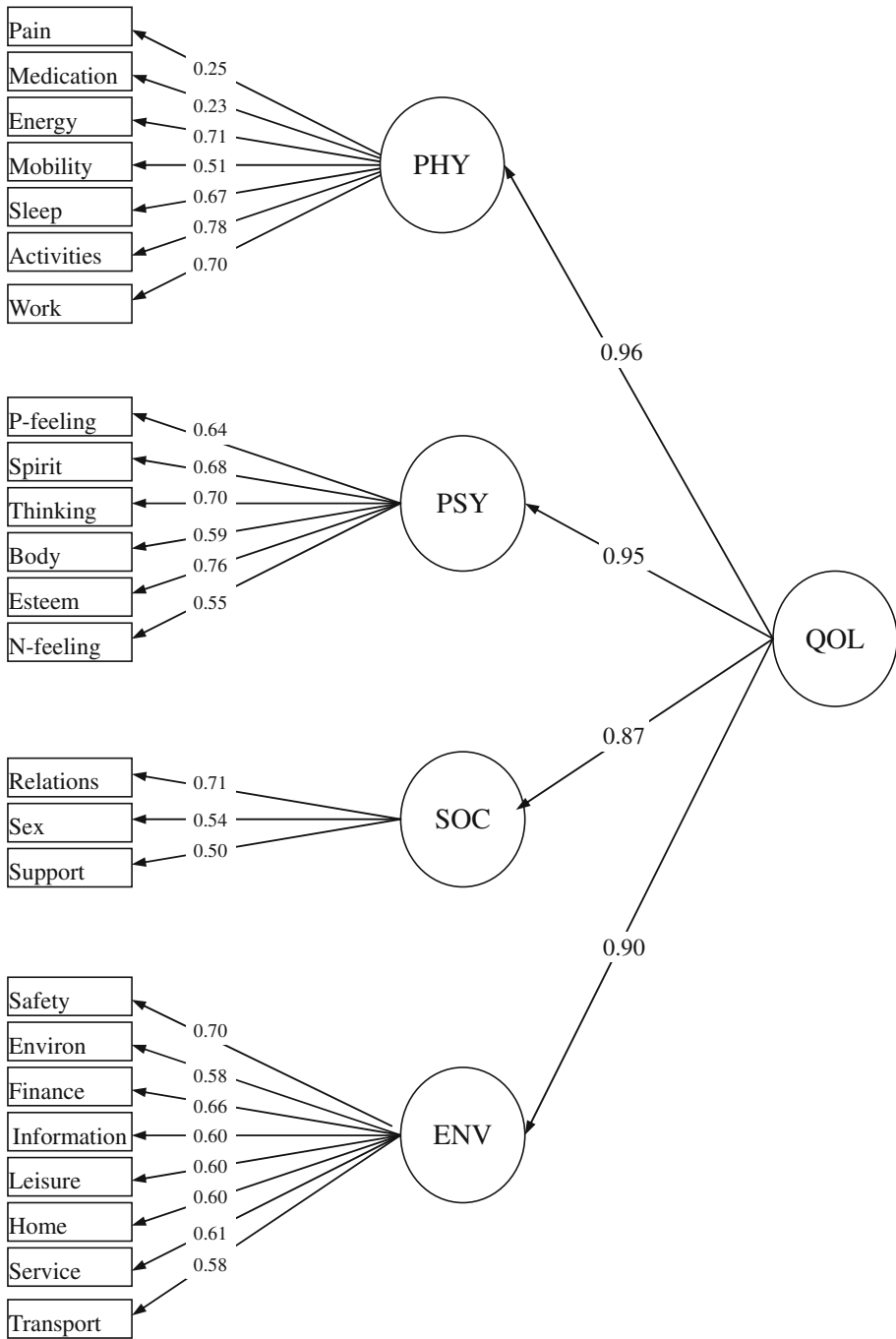
The hierarchical model for the web version of the WHOQOL-BREF was examined in this section. Figure 1 presents standardized estimates of the model with the web version of the

Table 5 Concurrent validity of the web version of the WHOQOL-BREF: standardized coefficients (β) for the multiple regressions on G1, G2, and G1 + G2 of the four domain scores

Dependent variable/predictors	PHY	PSY	SOC	ENV	F (4, 1,011)	R^2
G1 (Global QOL)	0.10	0.22	0.12	0.33	193.11*	0.43
G2 (General health)	0.10	0.17	0.11	0.33	147.12*	0.37
G1 + G2 (Sum of G1 and G2)	0.11	0.22	0.13	0.37	257.57*	0.51

PHY physical domain; *PSY* psychological domain; *SOC* social domain; *ENV* environmental domain; * $p < 0.001$

All β values were statistically significant at $p < 0.05$

Fig. 1 Results of confirmatory factor analysis (CFA) model for the web version of the WHOQOL-BREF. QOL, quality of life; PHY, physical domain; PSY, psychological domain; SOC, social domain; ENV, environmental domain; N-feeling, negative feeling; P-feeling, positive feeling. All estimates (standardized) were significant at $p < 0.01$

WHOQOL-BREF. For simplicity, the error variances of the observed and latent variables are not displayed. The parameters in the model were all significant at $p < 0.01$. Although the chi-square test rejected the model (χ^2 (248) = 1,162.33, $p < 0.01$), the values of the fit indices suggested that this model was retainable (NNFI = 0.97; CFI = 0.97; RMSEA = 0.062; 90% C.I. = 0.059–0.066; SRMR = 0.043).

Although a single factor model (all items were influenced by one factor) was also had similar result in model fit (χ^2 (252) = 1,421.16, $p < 0.01$; NNFI = 0.96; CFI = 0.96; RMSEA = 0.072; 90% C.I. = 0.069–0.075; SRMR = 0.048), the hierarchical model was better than the single factor model because values of fit index for model comparison of the hierarchical model (AIC = 1,320.78, CAIC = 1,628.81 and ECVI = 1.30) were lower than those values of the single factor model (AIC = 1,671.95, CAIC = 1,956.28 and ECVI = 1.65).

4 Discussion

This paper is the first one in the world to compare the web version WHOQOL-BREF with paper version, and to use the data collected from web to test its reliability and validity. The correlation coefficients at the domain level between the two versions, the test–retest reliability, mean difference tests, internal consistency, and concurrent validity were all acceptable, except that the psychological domain score became higher when web version was retested after 2 weeks. CFA confirmed the construct validity of the WHOQOL-BREF. All these findings suggest that the web version is as good as the paper version for evaluating HR-QOL. This result corroborates previous studies that have shown that web-based surveys of psychological distress and health status are equivalent to the traditional paper and pencil formats (Lin 2003; Epstein et al. 2001; Herrero and Meneses 2006; Knapp and Kirk 2003; Mangunkusumo et al. 2005; Pouwer et al. 1998).

In Study 1, the demographics of the two groups were not significantly different, so it is unlikely that the results were affected by these factors. All the correlation coefficients at the domain level were greater than 0.70, which is considered acceptable. The correlation was not affected by the different sequence of paper versus web formats; there was no statistical difference in the results between these two groups according to the z test. Moreover, mean difference tests in domain scores between web and paper versions were not significant. The test–retest reliabilities (ICC scores) were all over 0.75, confirming the web version of the WHOQOL-BREF in all four domains. In addition, using Bland and Altman (1986) method, the agreements between paper and web version were also satisfactory. It is reasonable that the lowest value occurred in the psychological domain, 0.79, because psychological feelings more easily fluctuate in a 2 week test–retest period than the other three domains, which are believed relatively more static. Even though the ICC score of the psychological domain was the lowest, its value is still acceptable by statisticians. However, in the test–retest scores of web version, the psychological domain score became higher when it was retested after 2 weeks.

In Study 2, Chronbach's α values were acceptable (>0.70) (Nunnally 1978), except for the social domain, which was 0.60. As the calculation of internal consistency is sensitive to the number of items, this lower value was expected because there are only three items in this domain. In addition, compared to other studies, in which the α values ranged from 0.51 to 0.77 (Yao et al. 2002), this value is also acceptable. The concurrent validity results in this study differed from previous research. One study found that for the general population in the US the best predictor of G1 was the psychological domain, and

for both G2 and G1 + G2 the physical domain was the best predictor (Yao et al. 2002). Other studies have found the best predictor of G1 to be the environmental domain, and both G2 and G1 + G2 were best predicted by the physical domain in the general population and for elderly people in Taiwan (Lin 2003; Lai et al. 2005). However, in this paper we found the environmental domain was the best predictor for G1, G2, and G1 + G2. This may reflect the fact that the people who participated in Study 2 were most concerned with environmental issues, which become somewhat unstable due to the political antagonism and economic stasis during the period of this study in Taiwan. Another possible reasons could be that they were younger (about two-thirds were under 40); most (over 80%) had no physical illness and less bothered by physical problems. Our finding that all β were significant corroborates previous studies (Lin et al. 2003; Ryan et al. 2002; Skevington et al. 2004) and implies that the measurement of HR-QOL and health should include all of these domains. To test whether the four domains fit the data, a second-order factor structure was also conducted. CFA was performed on the four factors using their corresponding indicators as a whole QOL model. The value of CFI indicates this model is appropriate for the data. All the results of the CFA were quite similar to those for the paper version of the WHOQOL-BREF (Yao et al. 2002).

However, it might be arguable that the four domains cannot exactly measure HR-QOL because the four domains only account for 37–51% variances of global QOL and health measures. This result was related to the measurement approach adopted by the WHO-QOL-BREF and the global measures of QOL and health. Specifically, the WHOQOL-BREF is a domain-measure instrument of health-related QOL, which is different form the global measures of QOL and health. According to Wu and Yao's (2007) study, both domain and global measures of QOL do assess the same latent construct of QOL, but because of the different measurement approach (domain vs. global), domain and global measures of QOL still have their unique, unshared variances. As a result, because there exits a difference in measuring approach between domain and global measures of QOL, we did not expect that a domain measure of QOL cannot fully explain the variance of a global QOL measure. And this is also a reason to explain why the four domain scores in the WHOQOL-BREF only can account for 43–51% variance of the global QOL and health.

As retaking a test the same day is too short an interval to avoid a recall bias (Ryan et al. 2002), and as we also needed to directly compare the two different versions as soon as possible, we used a 2 or 3 day interval for the comparison of the web and paper versions. In addition, to reduce the recall bias we did not inform participants that they would be asked to complete the questionnaire again later when they completed it the first time. Regarding the test–retest study, supposed that the HR-QOL of participants remain stationary within 2 weeks, this time interval seems to be long enough to check the reliability of web version WHOQOL.

Two types of missing data are discussed in HR-QOL measurement: one is unit non-response (the whole questionnaire is unanswered) and the other is item non-response (Curran et al. 1998). The first type occurred with both the web and paper formats, but the second type occurred only with the paper format in this study. We randomly asked five out of the eight non-respondents about the reason why they did not respond or fill either forms. Two reported they were too busy to do so, another two said they forgot, and the last one alleged she was off during the period when she was asked to fill out the questionnaire. All of them said that the design of requiring every question completed is acceptable. Thus, we believed that their reasons of non-response did not seem to affect the results of this study. The most frequent item non-response occurred with Item 21, which asks about sexual

satisfaction. This finding is the same as in previous studies (Chen et al. 2006; Hwang et al. 2003; Lin 2006) and can be explained by Taiwanese culture. Taiwanese people usually regard sex as a very private matter and feel uncomfortable to share with others. The other possible reason is that as over 60% were not married and they did not know how to evaluate their sexual life. This finding suggests that to minimize missing data for this item, greater elaboration of this question is needed. Another finding was that Item 21 had missing data in Group 2, but not in Group 1. It is possible that since the nursing staff in Group 1 were forced to complete the web version first, they then recalled their answer for the paper version 2 or 3 days later.

Some disadvantages of use of the internet to collect data have been pointed out, such as limited access to the internet and inability to use a computer, especially for minorities and people of low socio-economic status and limited education (Kalichman et al. 2002). In this study, because every psychiatric ward had one or more computers and the staff often used these computers for routine hospital work, access and inability were not problems. Even though this study did not ask participants which version they favored, studies have indicated many benefits of computer-based surveys, including elimination of missing data and manual errors in the coding procedure, and preference by participants (Pouwer et al. 1998; Wright et al. 1998). In addition, it seems to be more environmentally friendly and cost-efficient to use the web-format or computer-based researches, as it consumed only 480 pieces of paper in this study, but would use over 7,000 pieces of paper in paper-and-pencil mode.

The main limitations of this study are as follows: first, because it is difficult to establish a random sample from Taiwanese population to complete the web questionnaire and then 2 or 3 days later follow up with the paper format, the samples and the websites used in Study 1 and Study 2 were different. In addition, in Study 1, the sample size was small, mostly female with a homogenous job, and higher education than the general population. A recalculation of the power requirements showed that about 100 cases in study 1 would be required to raise the power of the study over 0.70. The results of the comparison might be different from a comparison with the general population. Second, in Study 2, even though the sample size of the participants was larger, they were still mostly younger females with a higher education level than the general population. Further, they were probably mostly regular Internet users. As a result, the reliability and validity test might be different from that of a sample of the general population. Although some studies have mentioned that computer literacy is not associated with the ability to complete a computer-based questionnaire (Bliven et al. 2001), it is still hard to say whether the results can be generalized to the general population of Taiwan, especially to non-computer users. In addition, the scale itself in WHOQOL-BREF might have some deficits such as some items did not show the adequate content validity in a study (Yao et al. 2008), and we did not explored this issue in this study. Despite of above limitations, this study still did provide evidence to show the web form WHOQOL-BREF is acceptable to evaluate the Health Related QOL.

Finally, we only examine the validity of the web version of the WHOQOL-BREF with standard items. In fact, The WHOQOL gruop (1995) encourages researchers to added national- or cultural-relevant items into the WHOQOL instrument. For example, in WHOQOL Taiwan version, two additional facets (eat and face) for Chinese culture were added. Even within the same Chinese culture, Taiwan, China, and Hong Kong versions of the WHOQOL-BREF added different items which are specific to their own areas (Yao and Wu in press). Thus, the WHOQOL instrument has its flexibility to face the demand of cultural relevance of the instrument. This kind of flexibility of the WHOQOL instrument

can compensate the paucity of Chinese measures on QOL as mentioned in Shek et al. (2005a, b), article and the special issue they edited. For example, Shek et al. (2005a, b) indicated that "while happiness and satisfaction are important components of QOL in the American culture, Chinese people emphasize forbearance, endurance and contending mentality (p2)", but items measuring forbearance, endurance and contending mentality are rarely included in commonly used QOL instrument. Thus, if one would like to assess the Chinese-specific construct of QOL, it is not available for him/her to find a useful instrument. Therefore, the flexibility of the WHOQOL instrument in adding national- or cultural-relevant items can facilitate us to develop a cultural-relevant QOL instrument within an international project. But in this study, we only examine the web version of the WHOQOL-BREF without considering the cultural-relevant issue here.

5 Conclusion

The results of the direct comparison between the two versions, and the reliability and validity tests were acceptable; this study provides the first empirical evidence in the world that the web version of the WHOQOL-BREF can be the alternative to the paper version. It may if the population under consideration is comfortable with the Internet for communication and access to the Internet is readily available. In addition, because hardcopies and manual coding are unnecessary for the web version, it seems to be a more effective and environmentally friendly option than the paper version.

Appendix

See Tables 6, 7 and 8.

Table 6 Correlation matrix of the four domains scores across web and paper versions for Group 1 in Study 1 ($n = 34$)

Domain	W-PHY	W-PSY	W-SOC	W-ENV	P-PHY	P-PSY	P-SOC
W-PHY	–						
W-PSY	0.49	–					
W-SOC	0.46	0.53	–				
W-ENV	0.72	0.66	0.57	–			
P-PHY	0.83	0.25	0.32	0.47	–		
P-PSY	0.60	0.79	0.46	0.58	0.57	–	
P-SOC	0.65	0.48	0.71	0.63	0.60	0.68	–
P-ENV	0.68	0.51	0.33	0.75	0.54	0.68	0.57

W web; *P* paper; *PHY* physical domain; *PSY* psychological domain; *SOC* social domain; *ENV* environmental domain

Table 7 Correlation matrix of the four domains scores across web and paper versions for Group 2 in Study 1 ($n = 38$)

Domain	P-PHY	P-PSY	P-SOC	P-ENV	W-PHY	W-PSY	W-SOC
P-PHY	–						
P-PSY	0.68	–					
P-SOC	0.58	0.70	–				
P-ENV	0.70	0.81	0.58	–			
W-PHY	0.77	0.67	0.54	0.74	–		
W-PSY	0.62	0.85	0.67	0.73	0.75	–	
W-SOC	0.47	0.63	0.76	0.49	0.61	0.76	–
W-ENV	0.55	0.73	0.49	0.83	0.75	0.75	0.63

W web; *P* paper; *PHY* physical domain; *PSY* psychological domain; *SOC* social domain; *ENV* environmental domain

Table 8 Correlation matrix of the four domains scores across T1 and T2 for Group 3 in Study 1 ($n = 60$)

Domain	T1-PHY	T1-PSY	T1-SOC	T1-ENV	T2-PHY	T2-PSY	T2-SOC
T1-PHY	–						
T1-PSY	0.59	–					
T1-SOC	0.49	0.53	–				
T1-ENV	0.66	0.77	0.65	–			
T2-PHY	0.65	0.47	0.39	0.53	–		
T2-PSY	0.44	0.84	0.45	0.66	0.57	–	
T2-SOC	0.32	0.46	0.83	0.51	0.50	0.52	–
T3-ENV	0.56	0.59	0.52	0.80	0.61	0.65	0.56

W web; *P* paper; *PHY* physical domain; *PSY* psychological domain; *SOC* social domain; *ENV* environmental domain

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246. doi:10.1037/0033-2909.107.2.238.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*, 307–310.

Bliven, B. D., Kaufman, S. E., & Spertus, J. A. (2001). Electronic collection of health-related quality of life data: Validity, time benefits, and patient preference. *Quality of Life Research, 10*, 15–22. doi: 10.1023/A:1016740312904.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Beverly Hills: Sage.

Bush, N., Donaldson, G., Moinpour, C., Haberman, M., Milliken, D., Markle, V., et al. (2005). Development, feasibility and compliance of a web-based system for very frequent QOL and symptom home self-assessment after hematopoietic stem cell transplantation. *Quality of Life Research, 14*, 77–93. doi: 10.1007/s11136-004-2394-2.

Bushnell, D. M., Reilly, M. C., Galani, C., Martin, M. L., Ricci, J. F., Patrick, D. L., et al. (2006). Validation of electronic data capture (EDC) of the irritable bowel syndrome-quality of life measure (IBS-QOL), the work productivity and activity impairment questionnaire for irritable bowel syndrome (WPAI:IBS) and the EuroQol (EQ-5D). *Value in Health, 9*, 98–105. doi:10.1111/j.1524-4733.2006.00087.x.

Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Ost, L. G., et al. (2007). Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior, 23*, 1421–1434. doi:10.1016/j.chb.2005.05.002.

Chen, K. H., Wu, C. H., & Yao, G. (2006). Applicability of the WHOQOL-BREF on early adolescence. *Social Indicators Research, 79*, 215–234. doi:10.1007/s11205-005-0211-0.

Curran, D., Molenberghs, D., Fayers, P., & Machin, D. (1998). Aspects of incomplete quality of life data in randomized trials: II. Missing data forms. *Statistics in Medicine, 17*, 697–709. doi:10.1002/(SICI)1097-0258(19980315/15)17:5/7<697::AID-SIM815>3.0.CO;2-Y.

Epstein, J., Klinkenberg, W. D., Wiley, D., & McKinley, L. (2001). Insuring sample equivalence across Internet and paper-and-pencil assessments. *Computers in Human Behavior, 17*, 339–346. doi:10.1016/S0747-5632(01)00002-4.

Fang, C. T., Hsiung, P. C., Yu, C. F., Chen, M. Y., & Wang, J. D. (2002). Validation of the World Health Organization quality of life instrument in patients with HIV infection. *Quality of Life Research, 11*, 753–762. doi:10.1023/A:1020870402019.

Fellinger, J., Holzinger, D., Dobner, U., Gerich, J., Lehner, R., Lenz, G., et al. (2005). An innovative and reliable way of measuring health-related quality of life and mental distress in the deaf community. *Social Psychiatry and Psychiatric Epidemiology, 40*, 245–250. doi:10.1007/s00127-005-0862-9.

Gordon, A. S., Panahian-Jand, M., Mccomb, F., Melegari, C., & Sharp, S. (2003). Characteristics of women with vulvar pain disorders: Responses to a Web-based survey. *Journal of Sex and Marital Therapy, 29*(Suppl 1), 45–58. doi:10.1080/713847126.

Herrero, J., & Meneses, J. (2006). Short Web-based versions of the perceived stress (PSS) and Center for Epidemiological Studies-Depression (CESD) scales: A comparison to pencil and paper responses among Internet users. *Computers in Human Behavior, 22*, 830–846. doi:10.1016/j.chb.2004.03.007.

Hoyle, R. H. (1995). The structural equation modeling approach: basic concepts and fundamental issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 1–15). Thousand Oaks: Sage.

Hsiung, P. C., Fang, C. T., Chang, Y. Y., Chen, M. Y., & Wang, J. D. (2005). Comparison of WHOQOL-BREF and SF-36 in patients with HIV infection. *Quality of Life Research, 14*, 141–150. doi:10.1007/s11136-004-6252-z.

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.

Hwang, H. F., Liang, W. M., Chiu, Y. N., & Lin, M. R. (2003). Suitability of the WHOQOL-BREF for community-dwelling older people in Taiwan. *Age and Ageing, 32*, 593–600. doi:10.1093/ageing/afg102.

Institute for Strategy and Competitiveness, Harvard Business School (2007). Value-based competition. http://www.hbs.edu/rhc/value.html. Retrieved 23 July 2007.

International Society for Quality of Life Research (2007). Are patient reported outcomes ready for prime time? http://www.isoqol.org/. Retrieved 15 June 2007.

Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8: structural equation modeling with the SIMPLIS command language*. Hillsdale: Lawrence Erlbaum Associates.

Kalichman, S. C., Weinhardt, L., Benotsch, E., & Cherry, C. (2002). Closing the digital divide in HIV/AIDS care: Development of a theory-based intervention to increase internet access. *AIDS Care, 14*, 523–537.

Kimura, T., Uchida, S., Tsuda, Y., & Eboshida, A. (2005). Computer-assisted measurement of perceived stress: An application for a community-based survey. *Hiroshima Journal of Medical Sciences, 54*, 61–65.

Knapp, H., & Kirk, S. A. (2003). Using pencil and paper, Internet and touch-tone phones for self-administered surveys: does methodology matter? *Computers in Human Behavior, 19*, 117–134. doi:10.1016/S0747-5632(02)00008-0.

Lai, J. N., Chen, H. J., Chen, C. M., Chen, P. C., & Wang, J. D. (2006). Quality of life and climacteric complaints amongst women seeking medical advice in Taiwan: Assessment using the WHOQOL-BREF questionnaire. *Climacteric, 9*, 119–128. doi:10.1080/13697130600635292.

Lai, K. L., Tzeng, R. J., Wang, B. L., Lee, H. S., Amidon, R. L., & Kao, S. (2005). Health-related quality of life and health utility for the institutional elderly in Taiwan. *Quality of Life Research, 14*, 1169–1180. doi:10.1007/s11136-004-3061-3.

Lin, C. H. (2003). Reliability of information provided by patients of a visual psychiatric clinic. *Psychiatric Services (Washington, DC), 54*, 1167–1168. doi:10.1176/appi.ps.54.8.1167.

Lin, T. H. (2006). Missing data imputation in quality-of-life assessment: Imputation for WHOQOL-BREF. *PharmacoEconomics, 24*, 917–925. doi:10.2165/00019053-200624090-00008.

Lin, T. H., Chang, H. Y., & Weng, W. S. (2003). The national health interview survey information: An overview. *Journal of Taiwan Public Health, 22*, 431–440.

Mangunkusumo, R. T., Moorman, P. W., Van Den Berg-de Ruiter, A. E., Van Der Lei, J., De Koning, H. J., & Raat, H. (2005). Internet-administered adolescent health questionnaires compared with a paper version in a randomized study. *Journal of Adolescent Health, 36*, 70.e1–70.e6.

Mason, V. L., Skevington, S. M., & Osborn, M. (2004). Development of a pain and discomfort module for use with the WHOQOL-100. *Quality of Life Research, 13*, 1139–1152. doi:10.1023/B:QURE.0000031344.53009.eb.

McKenna, S. P., & Doward, L. C. (2004). Integrating patient-reported outcomes. *Value in Health, 7*, S9–S12. doi:10.1111/j.1524-4733.2004.7s103.x.

Mullen, K. H., Berry, D. L., & Zierler, B. K. (2004). Computerized symptom and quality-of-life assessment for patients with cancer part II: Acceptability and usability. *Oncology Nursing Forum, 31*, E84–E89. doi:10.1188/04.ONF.E84-E89.

Naumann, V. J., & Byrne, G. J. (2004). WHOQOL-BREF as a measure of quality of life in older patients with depression. *International Psychogeriatrics, 16*, 159–173. doi:10.1017/S1041610204000109.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Porter, M. E., & Teisberg, E. O. (2006). *Redefining health care: Creating value-base competition on results*. Boston: Harvard Business School.

Pouwer, F., Snoek, F. J., van der Ploeg, H. M., Heine, R. J., & Brand, A. N. (1998). A comparison of the standard and the computerized versions of the well-being questionnaire (WBQ) and the diabetes treatment satisfaction questionnaire (DTSQ). *Quality of Life Research, 7*, 33–38. doi:10.1023/A:1008832821181.

Ryan, J. M., Corry, J. R., Attewell, R., & Smithson, M. J. (2002). A comparison of an electronic version of the SF-36 general health questionnaire to the standard paper version. *Quality of Life Research, 11*, 19–26. doi:10.1023/A:1014415709997.

Shek, D. T. L., Chan, Y. K., & Lee, P. S. N. (2005a). Quality of life in the global context: A Chinese response. *Social Indicators Research, 71*, 1–10. doi:10.1007/s11205-004-8035-x.

Shek, D. T. L., Chan Y. K., & Lee, P. S. N. (Eds.) (2005b). Special Issue on "Quality of life research in Chinese, Western and global contexts". *Social Indicators Research*, *71*, 1–539. doi:10.1007/s11205-004-8035-x.

Skevington, S. M., Lotfy, M. O'Connell, K. A. & WHOQOL Group. (2004). The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. A report from the WHOQOL group. *Quality of Life Research*, *13*, 299–310 (2004). doi:10.1023/B:QURE.0000018486.91360.00.

Taylor, W. J., Myers, J., Simpson, R. T., McPherson, K. M., & Weatherall, M. (2004). Quality of life of people with rheumatoid arthritis as measured by the World Health Organization Quality of Life instrument, short form (WHOQOL-BREF): Score distributions and psychometric properties. *Arthritis and Rheumatism, 51*, 350–357. doi:10.1002/art.20398.

The International Society for Pharmacoeconomics and Outcomes Research (2007). PRO task force: Changing mode of administration of instruments/ePRO. http://ispor.org/TaskForce/ePROTF.asp. Retrieved 17 May 2007.

The WHOQOL group. (1993). Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (the WHOQOL). *Quality of Life Research, 2*, 153–159. doi:10.1007/BF00435734.

The WHOQOL gruop. (1995). The World Health Organization Quality of Life assessment (WHOQOL): Position paper from the World Health Organization. *Social Science and Medicine, 41*, 1403–1409. doi:10.1016/0277-9536(95)00112-K.

The WHOQOL gruop. (1998). The World Health Organization Quality of Life assessment (WHOQOL): Development and general psychometric properties. *Social Science and Medicine, 46*, 1569–1585. doi:10.1016/S0277-9536(98)00009-4.

US Department of Health and Human Services, Food and Drug Administration (2007). Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims. http://www.fda.gov/CDER/GUIDANCE/5460dft.pdf. Retrieved 17 May 2007.

Wang, Y. C., Lee, C. M., Lew-Ting, C. Y., Hsiao, C. K., Chen, D. R., & Chen, W. J. (2005). Survey of substance use among high school students in Taipei: Web-based questionnaire versus paper-and-pencil questionnaire. *The Journal of Adolescent Health, 37*, 289–295. doi:10.1016/j.jadohealth.2005.03.017.

World Health Organization. (1996). *WHOQOL-BREF: Introduction, administration, scoring and generic version of the assessment—field trial version*. Geneva: WHO.

Wright, D. L., Aquilino, W. S., & Supple, A. J. (1998). A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly, 62*, 331–353. doi:10.1086/297849.

Wu, C. H., & Yao, G. (2007). Examining the relationship between global and domain measures of quality of life by three factor structure models. *Social Indicators Research, 84*, 189–202. doi:10.1007/s11205-006-9082-2.

Yang, S. C., Kuo, P. W., Wang, J. D., Lin, M. I., & Su, S. (2005). Quality of life and its determinants of hemodialysis patients in Taiwan measured by WHOQOL-BREF(TW). *American Journal of Kidney Diseases, 46*, 635–641. doi:10.1053/j.ajkd.2005.06.015.

Yao, G., Chung, C. W., Yu, C. F., & Wang, J. D. (2002). Development and verification of validity and reliability of the WHOQOL-BREF Taiwan version. *Journal of the Formosan Medical Association, 101*, 342–351.

Yao, G., Wu, C. H., & Yang, C. T. (2008). Examining the content validity of the WHOQOL-BREF from respondents' perspective by quantitative methods. *Social Indicators Research, 85*, 483–498. doi:10.1007/s11205-007-9112-8.

Yao, G., & Wu, C. H. Similarities and differences among three Chinese versions of the WHOQOL-Questionnaire: Comparisons of Taiwan version to the China and Hong-Kong versions. *Social Indicators Research*, in press.