Examining the content validity of the WHOQOL-BREF from respondents' perspective by quantitative methods

Grace Yao · Chia-huei Wu · Cheng-ta Yang

Received: 16 November 2006 / Accepted: 6 April 2007 / Published online: 17 May 2007
© Springer Science+Business Media B.V. 2007

Abstract Content validity, the extent to which a measurement reflects the specific intended domain of content, is a basic type of validity for a valid measurement. It was usually examined qualitatively and relied on experts' subjective judgments, not on respondents' responses. Therefore, the purpose of this study was to introduce and demonstrate how to use quantitative methods to examine the content validity for a certain measurement from respondents' viewpoint with quantitative methods. In the current study, content validity of the WHOQOL-BREF was examined with quantitative methods among 102 undergraduate students and 128 community adults. They were asked to rate the appropriateness of each item with the four domains of the WHOQOL-BREF respectively and also asked to sort the items into the four domains. Then, three quantitative methods on examining content validity were applied to analyze the rating and sorting data, including (1) proportion of substantive agreement, (2) substantive validity, and (3) ANOVA approach. These results were used to compare with the original content structure of the WHOQOL-BREF, to see if the original structure is consistent with the structure of subjects' judgments. The results showed that the content structure gained from appropriateness rating and item sorting was not totally consistent with the original content structure of the WHOQOL-BREF. Among 24 items, 12 items did not have adequate content validity. More discussion on these items and the issue of content validity were further provided.

Keywords Content validity · The WHOQOL-BREF

1 Introduction

Quality of life (QOL) instruments abound in the health care literature. Many of them were developed on the basis of psychological testing principles with solid psychometric

G. Yao (✉) · C.-h. Wu · C.-t. Yang
Department of Psychology, National Taiwan University, 1, Sec. 4, Roosevelt Road, Taipei, Taiwan
e-mail: kaiping@ntu.edu.tw

properties, such as internal consistency reliability, test–retest reliability, content validity, criterion-related validity and construct validity. Many of these properties were usually examined by existing numerical index or statistical methods, such as Cronbach’s alpha coefficients for internal consistency reliability, correlation analysis for test–retest reliability and criterion-related validity, and factor analysis for construct validity. In addition, many of these properties were usually examined directly from respondents’ responses.

However, among these psychometric properties, content validity, “the degree to which the questions, tasks or items on a test are representative of the universe of behavior the test was designed to sample (Gregory 2000)”, was usually examined qualitatively and relied on experts’ subjective judgments, not on respondents’ responses. This approach may have a drawback when experts’ subjective judgments were not representative to respondents’ ideas about the measurements. Therefore, Lennon (1956) proposed that content validity is rooted in the response of respondents. He indicated that content validity can be considered in a sense that “the extent to which a subject’s responses to the items of a test may be considered to be a representative sample of his responses to a real or hypothetical universe of situations which together constitute the area of concern to the person interpreting the test (p.295)”. Following this perspective, content validity should be examined from respondents’ viewpoint rather than experts’ judgments, because experts’ judgments was extremely based on their subjective knowledge and opinions. In line with this perspective, Anderson and Gerbing (1991) and Hinkin and Tracey (1999) proposed to use quantitative methods (their methods were introduced in Sect. “Method”) to analyze the judgments of a representative sample from population of interest for assessing content validity. They proposed different quantitative methods to collect and analyze respondents’ judgments on the content of the items in a certain measurement. Their methods not only adopted respondents’ perspective in examining content validity, but also provided a numerical or statistical result to evaluate the adequacy of content validity of each item. It is desirable to apply their methods to evaluate content validity for new developed measurements, especially for the subjective QOL measurements. Therefore, the main purpose of this study was to introduce and demonstrate how to use the methods proposed by Anderson and Gerbing (1991) and Hinkin and Tracey (1999) to examine the content validity for a certain measurement.

In the current study, the methods proposed by Anderson and Gerbing (1991) and Hinkin and Tracey (1999) to examine the content validity for the WHOQOL-BREF. The WHOQOL-BREF was developed from the WHOQOL-100, a cross-cultural QOL instrument developed by the World Health Organization (WHO) for assessing individuals’ subjective perception and feelings of life. The WHOQOL-100 contains 100 items for 25 facets (24 domain-specific facets and one general facet) covering six domains, including physical health, psychological state, level of independence, social relations, personal beliefs, and environment (The WHOQOL Group 1994). However, the WHOQOL-100 is too lengthy for some uses, for example, in large epidemiological studies where QOL is only one variable of interest or in clinical evaluations where patients did not have enough time or ability to complete all items. Thus, the WHOQOL-100 was simplified into a brief version, called the WHOQOL-BREF, by selecting 24 items from 24 facets (one item per facet) and two items from the general facet (Skevington et al. 2003). These 26 items covers four domains, including physical health, psychological state, social relations, and environment. These four domain scores were used to indicate an individual’s QOL (Skevington et al. 2003). Since then, the WHOQOL-BREF was commonly applied to academic research, clinical evaluation, cross-culture comparison, and so on.

To date, many psychometric and adaptation studies have showed that the WHOQOL-BREF has adequate psychometric properties among different populations (Berlim 2005; Izutsu et al. 2005; Skevington et al. 2003; Trompenaars et al. 2005; Yao et al. 2002). However, as we mentioned above, content validity on the WHOQOL-BREF was rarely examined from respondents' perspective with quantitative analysis. Therefore, the purpose of this study was to examine the content validity of the WHOQOL-BREF by quantitative methods proposed by Anderson and Gerbing (1991) and Hinkin and Tracey (1999). These results were used to compare with the original content structure of the WHOQOL-BREF, to see if the original structure is consistent with the structure of participants' judgments.

2 Method

2.1 Participants

One hundred and two National Taiwan University undergraduate students and 128 community adults at Taipei city participated in this study. For the student sample, 41 of them were male and 61 of the subjects were female. The range of age was from 18 to 25 with the mean of 20.63 (Std. = 1.43). For the community adult sample, 44 of them were male and 84 of the subjects were female. The range of age was from 19 to 63 with the mean of 39.26 (Std. = 10.50). These two samples were selected because the WHOQOL-BREF was a generic QOL instruments, student and adult populations are the applied targets of the WHOQOL-BREF. In addition, Yao et al.'s (2000) study revealed that student sample (homogeneous group) and adult sample (heterogeneous group) have different perceptions on scale descriptors selected for the WHOQOL-BREF. Although their study may not directly imply a difference between student sample and adult sample on content perception of items in the WHOQOL-BREF, however, their results recommend analyzing the judgments of these two groups separately. Therefore, both student and adult samples were used in the current study and their judgments of content validity were also compared to see if these two groups have different perceptions on items in the WHOQOL-BREF.

2.2 Instruments and procedure

2.2.1 *The standard WHOQOL-BREF*

In this study, the standard WHOQOL-BREF version was used. The 26 standard items were consisted of one item from each of the 24 facets of the WHOQOL-100 and two items from the overall QOL and general health facet. The standard WHOQOL-BREF version has been translated into Chinese for the Taiwan people (Yao et al. 2002). In Yao et al. (2002) study, exploratory and confirmatory factor analyses of the WHOQOL-BREF Taiwan version revealed a four-factor model (physical, psychological, social, and environmental domain factor). The internal consistency (Cronbach's alpha) coefficients ranged from 0.70 to 0.77 for the four domains (0.73 to 0.83 in this study). The test-retest reliability coefficients with interval 2–4 weeks ranged from 0.41 to 0.79 at item/facet level and 0.76 to 0.80 at domain level (all $p < 0.01$). Item-domain correlations were in the range of 0.53 to 0.78 for and inter-domain correlations were in the range of 0.51 to 0.64 for (all $p < 0.01$). In this study, only the 24 standard facet items of the WHOQOL-BREF were used.

2.2.2 Anderson and Gerbing's sorting method for the WHOQOL-BREF

Anderson and Gerbing (1991) method assumed that an item represents a single construct for respondents, so that they used sorting task to ask respondents to assign each item to a construct. Specifically, they asked participants to read each item and assign it to one construct or concept that, in their judgments, the item best indicated. We used the same procedure in this study. Participants were asked to sort the 24 facet items (randomly arranged on the answering sheet) in the WHOQOL-BREF into one of the four domains, including physical domain, psychological domain, social domain, and environmental domain.

Then, two indices proposed by Anderson and Gerbing (1991) were computed from the sorting data, including (1) proportion of substantive agreement (P_{SA}) and (2) substantive validity coefficient (C_{SV}). The proportion of substantive agreement (P_{SA}) is the proportion of respondents who assign an item to its intended construct, as follows:

$$P_{SA} = \frac{n_c}{N}$$

where n_c is the number of respondents assigning a measure to its posited construct and N is the total number of respondents. The possible range of P_{SA} is from 0.0 to 1.0. Higher value of P_{SA} indicates higher agreement of assignment for an item to its posited construct among respondents. The substantive validity coefficient (C_{SV}) reflects the extent to which respondents assign an item to its posited construct more than to any other construct. It is computed by

$$C_{SV} = \frac{n_c - n_o}{N},$$

where n_c is the number of respondents assigning a measure to its posited construct, n_o is the highest number of assignment of the item to any other construct, and N is the total number of respondents. The possible range of C_{SV} is from -1.0 to 1.0 . Positive value indicates assignments of an item to its posited construct are more than assignments to other construct. Conversely, negative value indicates assignments of an item to other construct are more than assignments to its posited construct. The absolute value indicates the strength of discrepancy between assignments to posited construct and assignments to other construct. However, there are no criterion values for P_{SA} and C_{SV} . Thus, in the current study, we used 0.30 as the cut point for P_{SA} . The 0.30 was used because there were four domains in our study. If one item was randomly assign to any domain, the expected proportion for P_{SA} is 0.25. Thus, we choose 0.30 as the criterion for P_{SA} , which was higher than 0.25. In addition, we also used 0.30 as the cut point for C_{SV} . Items with either P_{SA} or C_{SV} value below 0.30 were regarded as items with inadequate content validity.

2.2.3 Hinkin and Tracey's rating method for the WHOQOL-BREF

Based on the notion that any item may involve multiple constructs with one core construct and several minor constructs, Hinkin and Tracey (1999) proposed to use rating task to assess the strength of relations between an item and its constructs to see if an item has a strongest relation between the item and its posited construct. In their study, they asked respondents to rate the appropriateness of each item for different constructs on a 5-point Likert-type scale range from *not at all* (1) to *completely* (5). Then, a repeated one-way

ANOVA was applied to each item to see if an item has significant highest appropriateness score on its posited construct. Accordingly, in this study, participants were asked to rate the degree of adequacy of the 24 facet items on four domains, respectively. Specifically, participants were first asked to rate if each facet item belongs to the physical domain on a 7-point Likert-type scale ranging from 1 (extremely disagree) to 7 (extremely agree). And then, they were asked if the same items belong to psychological, social and environmental domains, respectively. Then, a repeated one-way ANOVA was applied to each item to see if each item has a highest score on its posited domain.

3 Results

3.1 Results on sorting data

The sorting data from both the student and adult samples were analyzed by Anderson and Gerbing (1991) methods, including (1) proportion of substantive agreement (P_{SA}) and (2) substantive validity coefficient (C_{SV}). Results for the student sample were presented in Table 1. In Table 1, the proportion of substantive agreement (P_{SA}) (proportion of assignment of an item in its posited domains) for each item was highlighted. The substantive validity coefficient (C_{SV}) for each item was displayed in the last column. According to the results, it is obvious that 10 items did not have acceptable values on these two indices (either P_{SA} or C_{SV} was lower than 0.30). These items were Item 6, 7, 10, 15, 16, 17, 19, 20, 21, and 23 (these item numbers were not the same as the item numbers in the standard WHOQOL-BREF). Two items were originally in the physical domain (Item 6 and 7), one item was originally in the psychological domain (Item 10), two items were originally in the social domain (Item 15 and 16), and five items were originally in the environmental domain (Item 17, 19, 20, 21 and 23).

The sorting results for the adult sample were presented in Table 2. Similarly, 11 items did not have acceptable values on these two indices (either P_{SA} or C_{SV} was lower than 0.30). They were Item 2, 6, 7, 14, 15, 16, 17, 19, 20, 21, and 23. Three items were originally in the physical domain (Item 2, 6, and 7), three items were originally in the social domain (Item 14, 15, and 16), and five items were originally in the environmental domain (Item 17, 19, 20, 21, and 23).

In addition, comparing the P_{SA} and C_{SV} values of acceptable items in Table 1 to those in Table 2, P_{SA} and C_{SV} values of acceptable items in Table 1 were generally higher than those in Table 2, revealing that students' judgments were closer to the original framework of the WHOQOL-BREF than adults' judgments.

3.2 Results on rating data

Further, the rating data was subjected to the repeated one-way ANOVA according to Hinkin and Tracey (1999) suggestion. Results for the student sample were presented in Table 3, in which the appropriateness ratings of each item on the four domains were shown, and the results of F -test were also displayed. The appropriateness value of the posited domains for each item was highlighted. After the F -test, pair comparison tests were conducted to indicate which domain has the statistically highest appropriateness value for each item. This result was shown in the last column of Table 3. Ideally, it is expected that

Table 1 Results of the proportion of substantive agreement (P_{SA}) and substantive-validity coefficient (C_{SV}) for the student sample

Original domain and item	Valid N	P_{SA}				C_{SV}
		PHY	PSY	SOC	ENV	
<i>Physical domain</i>						
1. To what extent do you feel that physical pain prevents you from doing what you need to do?	100	0.87	0.11	0.00	0.02	0.76
2. How much do you need any medical treatment to function in your daily life?	100	0.62	0.01	0.17	0.20	0.42
3. Do you have enough energy for everyday life?	100	0.87	0.12	0.01	0.00	0.75
4. How well are you able to get around?	100	0.75	0.06	0.04	0.15	0.60
5. How satisfied are you with your sleep?	100	0.79	0.13	0.00	0.08	0.66
6. How satisfied are you with your ability to perform your daily living activities?	100	0.36	0.51	0.11	0.02	0.15
7. How satisfied are you with your capacity for work?	100	0.07	0.43	0.44	0.06	0.37
<i>Psychological domain</i>						
8. How much do you enjoy life?	100	0.01	0.86	0.06	0.07	0.79
9. To what extent do you feel your life to be meaningful?	101	0.01	0.93	0.05	0.01	0.88
10. How well are you able to concentrate?	102	0.39	0.58	0.01	0.02	0.19
11. Are you able to accept your bodily appearance?	101	0.16	0.73	0.11	0.00	0.57
12. How satisfied are you with yourself?	101	0.00	0.97	0.02	0.01	0.95
13. How often do you have negative feelings, such as blue mood, despair, anxiety, depression?	101	0.02	0.98	0.00	0.00	0.97
<i>Social domain</i>						
14. How satisfied are you with your personal relationships?	101	0.00	0.26	0.70	0.04	0.45
15. How satisfied are you with your sex life?	100	0.40	0.41	0.19	0.00	0.22
16. How satisfied are you with the support you get from your friends?	101	0.00	0.40	0.51	0.09	0.11
<i>Environmental domain</i>						
17. How safe do you feel in your daily life?	101	0.01	0.32	0.29	0.38	0.06
18. How healthy is your physical environment?	100	0.01	0.01	0.01	0.97	0.96
19. Have you enough money to meet your needs?	100	0.06	0.09	0.33	0.52	0.19
20. How available to you is the information that you need in your day-to-day life?	101	0.02	0.02	0.36	0.60	0.24
21. To what extent do you have the opportunity for leisure activities?	100	0.16	0.14	0.29	0.41	0.12
22. How satisfied are you with the conditions of your living place?	101	0.00	0.03	0.01	0.95	0.92
23. How satisfied are you with your access to health services?	101	0.09	0.00	0.44	0.48	0.04
24. How satisfied are you with your transport?	101	0.00	0.06	0.11	0.83	0.72

PHY physical domain, *PSY* psychological domain, *SOC* social domain, *ENV* environmental domain

domain with the highest appropriateness value for each item is the posited domain of each item. However, according to Table 3, the highest domains for Item 6, 7, 10, 14, 15, 16, 17, 19, and 21 were not consistent to their posited domains. Some of them have a highest appropriateness value on one but not the posited domain, such as Item 7, 14, 15, and 16, and some of them have the same appropriateness values on two or three domains, such as Item 6, 10, 17, 19, and 21. Among these items, two items were originally in the physical

Table 2 Results of the proportion of substantive agreement (P_{SA}) and substantive-validity coefficient (C_{SV}) for the adult sample

Original domain and item	Valid <i>N</i>	<i>P</i> _{SA}				<i>C</i> _{SV}
		PHY	PSY	SOC	ENV	
<i>Physical domain</i>						
1. To what extent do you feel that physical pain prevents you from doing what you need to do?	121	0.77	0.17	0.01	0.05	0.60
2. How much do you need any medical treatment to function in your daily life?	118	0.43	0.09	0.31	0.17	0.13
3. Do you have enough energy for everyday life?	120	0.65	0.24	0.03	0.08	0.41
4. How well are you able to get around?	119	0.66	0.10	0.07	0.17	0.50
5. How satisfied are you with your sleep?	122	0.71	0.20	0.02	0.07	0.51
6. How satisfied are you with your ability to perform your daily living activities?	115	0.33	0.47	0.07	0.13	−0.14
7. How satisfied are you with your capacity for work?	120	0.08	0.54	0.22	0.16	−0.45
<i>Psychological domain</i>						
8. How much do you enjoy life?	119	0.07	0.71	0.11	0.11	0.58
9. To what extent do you feel your life to be meaningful?	120	0.08	0.80	0.11	0.01	0.69
10. How well are you able to concentrate?	120	0.28	0.68	0.01	0.03	0.39
11. Are you able to accept your bodily appearance?	119	0.18	0.79	0.01	0.02	0.60
12. How satisfied are you with yourself?	120	0.13	0.83	0.03	0.01	0.70
13. How often do you have negative feelings, such as blue mood, despair, anxiety, depression?	121	0.06	0.82	0.08	0.06	0.72
<i>Social domain</i>						
14. How satisfied are you with your personal relationships?	119	0.02	0.48	0.41	0.09	−0.06
15. How satisfied are you with your sex life?	110	0.63	0.36	0.00	0.01	−0.63
16. How satisfied are you with the support you get from your friends?	119	0.02	0.56	0.29	0.13	−0.28
<i>Environmental domain</i>						
17. How safe do you feel in your daily life?	118	0.02	0.37	0.36	0.25	−0.13
18. How healthy is your physical environment?	119	0.03	0.05	0.07	0.85	0.78
19. Have you enough money to meet your needs?	118	0.12	0.27	0.31	0.30	0.00
20. How available to you is the information that you need in your day-to-day life?	118	0.01	0.04	0.42	0.53	0.11
21. To what extent do you have the opportunity for leisure activities?	116	0.17	0.23	0.11	0.49	0.27
22. How satisfied are you with the conditions of your living place?	117	0.00	0.13	0.04	0.83	0.70
23. How satisfied are you with your access to health services?	119	0.13	0.09	0.54	0.24	−0.29
24. How satisfied are you with your transport?	121	0.03	0.07	0.27	0.63	0.36

PHY physical domain, *PSY* psychological domain, *SOC* social domain, *ENV* environmental domain

domain (Item 6 and 7), one item was originally in the psychological domain (Item 10), three items were originally in the social domain (Item 14, 15, and 16), and three items were originally in the environmental domain (Item 17, 19, and 21).

Similar findings were also observed for the adult sample in Table 4, in which Item 2, 6, 7, 14, 15, 16, 17, 19, 20, 21, and 23 were the items whose highest appropriateness values did not load on their posited domains. Some of them have a highest appropriateness value

Table 3 Results of appropriateness ratings of each item on four domains for the student sample

Original domain and item	Valid <i>N</i>	Mean value				<i>F</i> (<i>df</i> ₁ , <i>df</i> ₂)	Highest domain
		PHY	PSY	SOC	ENV		
<i>Physical domain</i>							
1. To what extent do you feel that physical pain prevents you from doing what you need to do?	101	5.50	3.95	2.62	2.59	<i>F</i> (3,300) = 134.26**	PHY
2. How much do you need any medical treatment to function in your daily life?	100	4.84	2.82	3.67	4.17	<i>F</i> (3,297) = 29.39**	PHY
3. Do you have enough energy for everyday life?	101	5.29	4.10	3.01	3.35	<i>F</i> (3,300) = 62.78**	PHY
4. How well are you able to get around?	101	5.36	3.40	3.65	4.35	<i>F</i> (3,300) = 42.38**	PHY
5. How satisfied are you with your sleep?	101	5.45	4.06	2.75	3.69	<i>F</i> (3,300) = 72.12**	PHY
6. How satisfied are you with your ability to perform your daily living activities?	101	4.77	4.82	3.66	3.61	<i>F</i> (3,300) = 27.28**	PHY, PSY
7. How satisfied are you with your capacity for work?	101	3.89	4.54	4.32	3.32	<i>F</i> (3,300) = 14.91**	PSY
<i>Psychological domain</i>							
8. How much do you enjoy life?	101	3.89	5.69	4.35	4.47	<i>F</i> (3,300) = 36.70**	PSY
9. To what extent do you feel your life to be meaningful?	101	3.44	6.04	3.86	2.82	<i>F</i> (3,300) = 102.98**	PSY
10. How well are you able to concentrate?	101	4.58	4.65	2.73	3.18	<i>F</i> (3,300) = 50.27**	PSY, PHY
11. Are you able to accept your bodily appearance?	101	4.25	5.25	3.81	2.53	<i>F</i> (3,300) = 65.98**	PSY
12. How satisfied are you with yourself?	101	3.89	5.91	3.66	2.69	<i>F</i> (3,300) = 105.16**	PSY
13. How often do you have negative feelings, such as blue mood, despair, anxiety, depression?	101	3.87	5.99	3.55	2.86	<i>F</i> (3,300) = 88.93**	PSY
<i>Social domain</i>							
14. How satisfied are you with your personal relationships?	101	3.35	5.49	5.09	3.59	<i>F</i> (3,300) = 50.96**	PSY
15. How satisfied are you with your sex life?	100	5.09	4.40	3.55	2.85	<i>F</i> (3,297) = 49.74**	PHY
16. How satisfied are you with the support you get from your friends?	101	3.30	5.52	4.90	3.53	<i>F</i> (3,297) = 53.63**	PSY

Table 3 continued

Original domain and item	Valid <i>N</i>	Mean value				<i>F</i> (<i>df</i> ₁ , <i>df</i> ₂)	Highest domain
		PHY	PSY	SOC	ENV		
<i>Environmental domain</i>							
17. How safe do you feel in your daily life?	101	3.79	5.40	5.01	5.31	<i>F</i> (3,297) = 32.12**	PSY, SOC, ENV
18. How healthy is your physical environment?	101	4.38	3.34	4.10	6.12	<i>F</i> (3,300) = 78.81**	ENV
19. Have you enough money to meet your needs?	101	3.63	3.47	4.08	4.34	<i>F</i> (3,300) = 7.66**	SOC, ENV
20. How available to you is the information that you need in your day-to-day life?	100	3.61	3.36	4.89	5.76	<i>F</i> (3,297) = 60.95**	ENV
21. To what extent do you have the opportunity for leisure activities?	101	4.48	4.00	4.04	4.46	<i>F</i> (3,300) = 3.63*	PHY, ENV
22. How satisfied are you with the conditions of your living place?	101	4.04	3.94	3.92	5.95	<i>F</i> (3,300) = 59.96**	ENV
23. How satisfied are you with your access to health services?	101	4.67	3.20	4.51	5.30	<i>F</i> (3,300) = 34.71**	ENV
24. How satisfied are you with your transport?	101	3.44	3.36	4.20	5.78	<i>F</i> (3,300) = 65.22**	ENV

PHY physical domain, PSY psychological domain, SOC social domain, ENV environmental domain

* $p < 0.05$, ** $p < 0.01$

on one but not the posited domain, such as Item 7, 14, 16, and 17, and some of them have the same appropriateness values on two or three, even four domains, such as Item 2, 6, 15, 19, 20, 21, and 23. Among these items, three items were originally in the physical domain (Item 2, 6, and 7), three items were originally in the social domain (Item 14, 15, and 16), and five items were originally in the environmental domain (Item 17, 19, 20, 21, and 23).

3.3 Comparison between sorting and rating results for each sample

In this section, the results of comparison between sorting and rating data on the student and adult samples were presented respectively. Table 5 presented (1) the assigned domain with highest proportion of assignment in the sorting task and (2) the domain with highest appropriateness values in the rating task for each item and for each sample respectively.

First, regarding the student sample, ten items, including Item 6, 7, 10, 14, 15, 16, 17, 19, 21, and 23, consistently showed worse results either on sorting or rating tasks or both. For these items, they were assigned to other domains rather than their posited domains in sorting task or/and they did not have the highest appropriateness values on their posited domains. Additionally, it should be noted that in Table 5, Item 20 did not have inconsistent results between sorting task and rating task and it indeed loaded on its posited domain (environmental domain), however, because Item 20 have lower substantive validity coefficients ($C_{SV} = 0.24$, indicating that it has highest proportion of assignment to the environmental domain, but this proportion is not substantively higher than the proportion of assignment to other domain when 0.30 was used as cut point), thus, it was also regarded as an item with inadequate content validity.

Further, regarding the adult sample, eleven items, including Item 2, 6, 7, 14, 15, 16, 17, 19, 20, 21, and 23, consistently showed worse results both on sorting and rating tasks. For these items, they were assigned to other domains than their posited domains or have a highest proportion of assignment to their posited domains, but the proportion is not substantively higher than proportion of assignment to other domain in the sorting task. In addition, they also did not have the highest appropriateness values on their posited domains. Generally, the findings from the sorting and rating tasks either for the student sample or for the adult sample were highly consistent. That is, items with worse results in sorting task usually have worse results in rating task as well.

3.4 Comparison between results of the student and adult samples

Taken all the results from the student and adult samples, it can be found that ten items exhibited inadequate content validity for both samples. These items were Item 6, 7, 14, 15, 16, 17, 19, 20, 21, and 23. In addition, Item 10 only showed a worse result for student sample, and Item 2 only showed worse results for adult sample. If all items with inadequate content validity were taken into account, then, there were 12 items having inadequate content validity.

Among these 12 items, three items were originally in the physical domain (Item 2, 6, and 7), one item was originally in the psychological domain (Item 10), three items were originally in the social domain (Item 14, 15, and 16), and five items were originally in the environmental domain (Item 17, 19, 20, 21, and 23). Generally, the inadequate items from physical domain (Item 2, 6, and 7) usually confounded with psychological domain, the inadequate item from psychological domain (Item 10) usually confounded with physical

Table 4 Results of appropriateness ratings of each item on four domains for the adult sample

Original domain and item	Valid N	Mean value			$F(df_1, df_2)$	Highest domain
		PHY	PSY	SOC		
<i>Physical domain</i>						
1. To what extent do you feel that physical pain prevents you from doing what you need to do?	116	4.84	4.09	3.21	3.27	$F(3,345) = 31.53^{**}$ PHY
2. How much do you need any medical treatment to function in your daily life?	120	3.89	2.85	4.04	3.75	$F(3,357) = 12.90^{**}$ PHY, SOC, ENV
3. Do you have enough energy for everyday life?	118	4.97	4.38	3.78	3.72	$F(3,351) = 19.80^{**}$ PHY
4. How well are you able to get around?	117	5.50	4.21	4.20	4.63	$F(3,348) = 18.60^{**}$ PHY
5. How satisfied are you with your sleep?	117	5.22	4.40	3.44	4.09	$F(3,348) = 30.29^{**}$ PHY
6. How satisfied are you with your ability to perform your daily living activities?	120	4.82	4.77	4.26	4.26	$F(3,357) = 5.92^{**}$ PHY, PSY
7. How satisfied are you with your capacity for work?	121	4.32	5.09	4.23	3.95	$F(3,360) = 14.39^{**}$ PSY
<i>Psychological domain</i>						
8. How much do you enjoy life?	119	4.30	5.38	4.07	4.39	$F(3,354) = 19.25^{**}$ PSY
9. To what extent do you feel your life to be meaningful?	119	4.23	5.70	4.22	3.73	$F(3,354) = 38.35^{**}$ PSY
10. How well are you able to concentrate?	118	4.88	5.22	3.37	3.69	$F(3,351) = 43.27^{**}$ PSY
11. Are you able to accept your bodily appearance?	119	4.55	5.40	3.82	3.59	$F(3,354) = 33.31^{**}$ PSY
12. How satisfied are you with yourself?	118	4.31	5.64	3.75	3.57	$F(3,351) = 43.96^{**}$ PSY
13. How often do you have negative feelings, such as blue mood, despair, anxiety, depression?	115	3.68	5.17	3.58	3.41	$F(3,342) = 32.37^{**}$ PSY
<i>Social domain</i>						
14. How satisfied are you with your personal relationships?	119	4.03	5.50	4.47	3.78	$F(3,354) = 30.03^{**}$ PSY
15. How satisfied are you with your sex life?	115	4.98	4.81	3.43	3.37	$F(3,342) = 39.73^{**}$ PHY, PSY
16. How satisfied are you with the support you get from your friends?	120	3.94	5.38	4.48	3.86	$F(3,357) = 26.50^{**}$ PSY

Table 4 Continued

Original domain and item	Valid <i>N</i>	Mean value			<i>F(df₁, df₂)</i>	Highest domain
		PHY	PSY	SOC		
<i>Environmental domain</i>						
17. How safe do you feel in your daily life?	120	3.97	5.29	4.88	4.96	PSY
18. How healthy is your physical environment?	117	4.32	3.86	4.49	5.38	ENV
19. Have you enough money to meet your needs?	120	3.85	4.15	4.20	4.04	No differences
20. How available to you is the information that you need in your day-to-day life?	120	4.05	4.18	5.18	5.06	SOC, ENV
21. To what extent do you have the opportunity for leisure activities?	120	4.65	4.56	4.64	4.80	No differences
22. How satisfied are you with the conditions of your living place?	117	4.18	4.23	4.52	5.31	ENV
23. How satisfied are you with your access to health services?	118	4.45	3.95	4.89	4.56	PHY, SOC, ENV
24. How satisfied are you with your transport?	119	4.03	4.12	4.86	5.21	ENV

PHY physical domain, PSY psychological domain, SOC social domain, ENV environmental domain

* $p < 0.05$, ** $p < 0.01$

Table 5 Highest proportion of substantive agreement domain and highest appropriateness domain of each item for the student and adult samples

Original domain and item	Student sample		Adult sample	
	Sorting	Rating	Sorting	Rating
<i>Physical domain</i>				
1. To what extent do you feel that physical pain prevents you from doing what you need to do?	PHY	PHY	PHY	PHY
2. How much do you need any medical treatment to function in your daily life?	PHY	PHY	PHY	PHY, SOC, ENV
3. Do you have enough energy for everyday life?	PHY	PHY	PHY	PHY
4. How well are you able to get around?	PHY	PHY	PHY	PHY
5. How satisfied are you with your sleep?	PHY	PHY	PHY	PHY
6. How satisfied are you with your ability to perform your daily living activities?	PSY	PHY, PSY	PSY	PHY, PSY
7. How satisfied are you with your capacity for work?	PSY, SOC	PSY	PSY	PSY
<i>Psychological domain</i>				
8. How much do you enjoy life?	PSY	PSY	PSY	PSY
9. To what extent do you feel your life to be meaningful?	PSY	PSY	PSY	PSY
10. How well are you able to concentrate?	PSY	PSY, PHY	PSY	PSY
11. Are you able to accept your bodily appearance?	PSY	PSY	PSY	PSY
12. How satisfied are you with yourself?	PSY	PSY	PSY	PSY
13. How often do you have negative feelings, such as blue mood, despair, anxiety, depression?	PSY	PSY	PSY	PSY
<i>Social domain</i>				
14. How satisfied are you with your personal relationships?	SOC	PSY	PSY, SOC	PSY
15. How satisfied are you with your sex life?	PHY, PSY	PHY	PHY	PHY, PSY
16. How satisfied are you with the support you get from your friends?	SOC	PSY	PSY	PSY
<i>Environmental domain</i>				
17. How safe do you feel in your daily life?	PSY, ENV, SOC	PSY, SOC, ENV	PSY, SOC	PSY
18. How healthy is your physical environment?	ENV	ENV	ENV	ENV
19. Have you enough money to meet your needs?	ENV	SOC, ENV	ENV, SOC, PSY	No differences
20. How available to you is the information that you need in your day-to-day life?	ENV	ENV	SOC, ENV	SOC, ENV
21. To what extent do you have the opportunity for leisure activities?	ENV	PHY, ENV	ENV	No differences
22. How satisfied are you with the conditions of your living place?	ENV	ENV	ENV	ENV
23. How satisfied are you with your access to health services?	ENV, SOC	ENV	SOC	PHY, SOC, ENV
24. How satisfied are you with your transport?	ENV	ENV	ENV	ENV

PHY physical domain, *PSY* psychological domain, *SOC* social domain, *ENV* environmental domain

domain, the inadequate items from social domain (Item 14, 15, and 16) usually confounded with psychological and physical domains, and the inadequate items from environmental domain (Item 17, 19, 20, 21, and 23) usually confounded with psychological and social domains. Among the four domains, social domain has the worst result. The all three items in the social domains have inadequate content validity for the student and adult samples. Generally, it is obvious that the findings from the student and adult samples were highly consistent.

4 Discussion

The purpose of this study was to examine the content validity of the WHOQOL-BREF using methods proposed by Anderson and Gerbing (1991) and Hinkin and Tracey (1999). Their approaches were based on the notion that content validity is rooted in the respondents from interested population, thus, they proposed to examine the content validity of items using a representative sample's judgments. In addition, in order to provide a convenient criterion in evaluation, they also proposed quantitative methods to analyze representative sample's judgments in sorting and rating tasks, respectively. In the current study, their approaches were applied to examine the content validity of items in the WHOQOL-BREF.

Generally, results showed that there were 12 items exhibiting inadequate content validity regarding to (1) proportion of substantive agreement (P_{SA}), (2) substantive validity (C_{SV}), and (3) repeated ANOVA approach. These 12 items confounded with the meanings of different domains. As mentioned in Sect. "Results", generally, the items from physical domain usually confounded with psychological domain, the items from psychological domain usually confounded with physical domain, the items from social domain usually confounded with psychological and physical domains, and the items from environmental domain usually confounded with psychological and social domains. This finding may imply that the 12 items are not good at representing their posited domains and may also result in inadequate discriminant validity of the four domains in the WHOQOL-BREF. That is, because these 12 items confounded with different domains, relations among the four domains based on the domains score computed from 24 facet items may be overestimated, such as the findings exhibited by the WHOQOL-Taiwan Group (2001) and Wang et al. (2006).

Indeed, for the WHOQOL-BREF-Taiwan version, several empirical studies showed that the four domain scores of the WHOQOL-BREF-Taiwan version were significantly correlated (e.g., 0.48–0.63, in the WHOQOL-Taiwan Group's study 2001). In addition, calculating from the confirmatory factor analysis model in Yao et al.'s (2002) study, correlations among the four-domain factors ranged from 0.77 to 0.91. Moreover, a result of multidimensional Rasch analysis also revealed that the correlations among latent factors of the four domains in the WHOQOL-BREF-Taiwan version were higher than 0.80 (Wang et al., 2006). Recently, Hsiao et al. (2005) conducted multitrait-multimethod (MTMM) design to examine the convergent and discriminant validity of the four domains of the WHOQOL-BREF for Taiwan people. Their results also showed that the four domains have strong correlations, even though method effects were controlled in the MTMM analysis. These empirical findings for the WHOQOL-BREF lead us to suspect that the four domains did not tap distinct constructs.

However, what causes this phenomenon? It would be possible that the high correlations among the four domains were due to the items confounded with different domains in the WHOQOL-BREF. If we select items that only tap a single domain, the high correlations

among the four domains may be eliminated. Nevertheless, it is also possible that the four domains in the WHOQOL-BREF were dependent in nature. That is, these four constructs are related inherently. Intuitively, this is highly possible, for example, satisfying with oneself (psychological domain in the WHOQOL) can be influenced by his/her capacity for work and personal relationships. Therefore, it is nature that items for different domains may have high correlations. Thus, no matter which items were selected, we would still find high correlations among these four domains. However, because there were items confounded with different domains in the WHOQOL-BREF, we did not know if the correlations among the four domains reflect the nature dependence of the four domains or the biased correlations due to several confounded items. Accordingly, we would like to emphasize the importance of content validity. It is obvious now that if we cannot ensure the content validity of each item for a certain measurement, we did not know how to interpret the results we gain, such as the higher correlations among the four domains of the WHOQOL-BREF as noted above.

Thus, we suggest WHOQOL instrument researchers and other QOL instrument researchers as well re-thinking the meanings of items from respondents' viewpoints. In addition, we would also like to recognize that the finding in the current study may be influenced by a cultural or national factor. Because we only conducted our study for Taiwan people, we did not know if these items have the same performance in other nations or cultures. It is possible that the standard items in the WHOQOL-BREF have excellent content validity based on the procedure of the current study in other cultures. Therefore, it is important for the WHOQOL researchers to ensure the content validity of the standard items within their cultures or nations. Because these standard items were used in cross-cultural comparison, it is worth showing that all these items can represent the same constructs for respondents from different cultural backgrounds. Moreover, it is also worth investigating content validity of the items in the WHOQOL-BREF on patients. Although this study showed that student and adult sample generally have similar perception on the standard items in the WHOQOL-BREF, however, it is unsure that whether patients with different diseases would have the same results if we applied the same procedure to assess the content validity of these standard items.

Generally, this study demonstrates (1) how to assess content validity from respondents' judgments and (2) examine the content validity of the WHOQOL-BREF for the Taiwan population by quantitative methods. In the future study, researchers may consider this respondent-centered and quantitative content validity when a measurement is developed and adapted.

Acknowledgment This study was supported by National Science Council (NSC 94-2413-H-002-018-NSC 95-2413-H-002-002) and National Health Research Institute (NHRI-EX95-9204PP, NHRI-EX96-9204PP)

References

- Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *The Journal of Applied Psychology*, 76, 732–740.
- Berlim, M. T., Pavanello, D. P., Caldieraro, M. A., & Fleck, M. P. (2005). Reliability and validity of the WHOQOL BREF in a sample of Brazilian outpatients with major depression. *Quality of Life Research*, 14, 561–564.
- Gregory, R. J. (2000). *Psychological testing: History, principles, and applications* (3rd ed.). Needham Heights: Allyn and Bacon.

- Hinkin, T. R., & Tracey, J. B. (1999). An analysis of variance approach to content validation. *Organizational Research Methods*, 2, 175–186.
- Hsiao, Y. Y., Wu, C. H., & Yao, G. (2005). *Examining the convergent and discriminant validity of the WHOQOL-BREF using the multitrait-multimethod (MTMM) approach*. Poster presented at 12th Annual Conference International Society for Quality of Life Research, San Francisco, CA, USA.
- Izutsu, T., Tsutsumi, A., Islam, A., Matsuo, Y., Yamada, H. S., Kurita, H., et al. (2005). Validity and reliability of the Bangla version of WHOQOL-BREF on an adolescent population in Bangladesh. *Quality of Life Research*, 14, 1783–1789.
- Lennon, R. T. (1956). Assumptions underlying the use of content validity. *Educational and Psychological Measurement*, 16, 294–304.
- Skevington, S. M., Lotfy, M., O'Connell, K., & WHOQOL Group. (2003). 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.
- The WHOQOL Group. (1994). The development of the World Health Organization Quality of Life Assessment Instrument (WHOQOL). In J. Orley, & W. Kuyken (Eds.), *Quality of life assessment: International perspectives* (pp. 41–57). Berlin: Springer.
- The WHOQOL-Taiwan Group. (2001). *The user's manual of the development of the WHOQOL-BREF Taiwan version (1st ed.)*. Taipei: Institute of Occupational Medicine and Industrial Hygiene, College of Public Health, National Taiwan University.
- Trompenaars, F. J., Masthoff, E. D., Van Heck G. L., Hodiament P. P., & De Vries J. (2005). Content validity, construct validity, and reliability of the WHOQOL-BREF in a population of Dutch adult psychiatric outpatients. *Quality of Life Research*, 14, 151–160.
- Wang, W. C., Yao, G., Tsai, Y. J., Wang, J. D., & Hsieh, C. L. (2006). Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. *Quality of Life Research*, 15, 607–620.
- Yao, G., Chung, C. W., Yu, C. F., & Wang J. D. (2002). Development and verification of reliability and validity of the WHOQOL-BREF Taiwan version. *Journal of the Formosan Medical Association*, 101, 342–351.
- Yao, G., Lin, M. R., & Wang, J. D. (2000). A comparative study on scale descriptor selection: Heterogeneous group vs. homogeneous group. *Chinese Journal of Psychology*, 41, 141–153.