Development of Short Versions for the WHOQOL-OLD Module

Jiqian Fang, PhD,1 Mick Power, PhD,2 Yueqing Lin, MSc,1 Jinxin Zhang, PhD,1 Yuantao Hao, PhD,1 and Somnath Chatterji, PhD3

1School of Public Health, Sun Yat-sen University, Guangzhou, China.
2Clinical Psychology, Medical School, Edinburgh University, Edinburgh, UK.
3World Health Organization, Geneva, Switzerland.

*Address correspondence to Jiqian Fang, School of Public Health, Sun Yat-sen University, 74 Zhongshan Road II, Guangzhou 510080, China.
E-mail: fangjq@mail.sysu.edu.cn

Received April 25, 2011; Accepted July 18, 2011
Decision Editor: Rachel Pruchno, PhD

Purpose of the study: To explore short-form versions of World Health Organization Quality of Life (WHOQOL-OLD) with acceptable psychometric properties, which was developed for older adults by the WHOQOL research group, containing 24 items initially. Design and Methods: We randomly sampled two-thirds of respondents from the data of WHOQOL-OLD field study (N = 5,566), as a developmental sample, and the remaining third as a validation sample. Three approaches (item response theory [IRT] and regression analysis [REG], classical test theory [CTT] and REG, and CTT and IRT and REG) were performed to develop three short-form scales with six items each using the developmental sample. The reliability and criterion validity of the three short-form scales were evaluated using the validation sample. Results: The version 1 of short-form scales contained items 6, 11, 12, 16, 20, and 21; version 2 contained items 2, 6, 11, 12, 17, and 22; and version 3 contained items 4, 6, 17, 19, 20, and 24. The correlation coefficient between the total scores of the three versions and WHOQOL-OLD were .918, .925, and .922, respectively. The internal consistency reliability coefficients of the three versions were .681, .678, and .649, respectively. Implications: The three versions of short-form WHOQOL-OLD contained the best items of the original module, much shorter, and with good internal consistency and criterion validity as a whole.

Key Words: Biostatistics, Regression models, Factor analysis, Disabilities, Quality of life, Measurement, Psychometrics, Quantitative research Methods, Evaluation

There are increasing concerns about aging populations worldwide; therefore, more researchers are interested in the evaluation of quality of life (QOL) and interventions to improve older people’s health, independence, activity, social, and economic participation (Peek, Ray, Patel, Stoebner-May, & Ottenbacher, 2004; Pruchno, Wilson-Genderson, Rose, & Cartwright, 2010; Schulz et al., 2010; Trigg, Skevington, & Jones, 2007). Since the early 1990s, the World Health Organization Quality of Life Group (WHOQOL Group) has developed two generic QOL instruments, the WHOQOL-100 (WHO, 1998) and its short form, the WHOQOL-BREF (WHOQOL Group, 1998). They include four domains: physical, psychological, social, and environmental aspects of QOL. Considering the applicability of these two instruments for older adults, the WHOQOL research group developed a WHOQOL-OLD module for older adults, containing six facets, with four items in each facet (Power, Quinn, Schmidt, & WHOQOL-OLD Group, 2005). The different language versions of WHOQOL-OLD have been evaluated and have shown good reliability and validity (Chachamovich, Fleck,
Trentini, & Power, 2008; Eser, Saatli, Eser, Baydur, & Fidaner, 2010; Fleck, Chachamovich, &Trentini, 2006; Halvorsrud & Kalfoss, 2008). However, some studies found some facets or items of it were poor. For example, it were found that “death and dying” and “sensory ability” facets showed inadequate fit to the model that was adequate for most items, and the item “problems with sense functioning affect ability to interact” in the sensory ability facet showed uniform differential item functioning (DIF) for age (Chachamovich et al., 2008); the sensory ability and “intimacy” facets might be less relevant for QOL in older adults (Fleck et al., 2006); and the WHOQOL-OLD had poor reliability and validity in an ethnically diverse sample (Bowling, 2009). In addition, the length of scale is an important influence factor for the quality of investigation. With the advance of QOL study, developing shortened versions and short forms of QOL assessments has become very significant. For example, the SF-36 has a series of different short forms including SF-12, SF-10, and SF-8, and the SCL-90-R varies with respect to number of items included and the dimensional structure identified from 90 items to 9, 14, or 27 items (three short-form versions). The WHOQOL-OLD with more items and needing more time to complete may be problematic for some older adults with low vision or physical disability or serious illness. Therefore, a brief tool that is easy to complete and interpret could be more readily incorporated into routine care or used in larger population studies.

This paper aims to develop one or more short versions of the WHOQOL-OLD module with good measurement properties and with one item per facet from the initial WHOQOL-OLD module of 24 items (Power et al., 2005).

**Methods**

**Data and Sample**

A secondary analysis was conducted based on the data from the WHOQOL-OLD field study, which has been reported in detail elsewhere (Power et al., 2005). The data set included N = 5,566 respondents coming from 20 international centers. We randomly sampled two-thirds of them as the developmental sample and the remaining third as a validation sample. The developmental sample was used to develop the short-form scale, whereas the validation sample was used to assess the reliability and validity of the possible scales.

The initial WHOQOL-OLD module covers six facets (sensory abilities, autonomy, past, present, and future activities, social participation, death and dying, and intimacy), of which each contains four items. The response format is a 5-point Likert scale (i.e., score 1 = not at all, 2 = a little, 3 = a moderate amount, 4 = very much, 5 = an extreme amount). Items OLD_01, OLD_02, and OLD_10 from the sensory abilities facet, and items OLD_06, OLD_07, and OLD_08, OLD_09 from the death and dying facet are reverse-coded items. Higher scores indicate higher QOL. The facets and scale scores were transformed to a 0–100 scale.

**Analytic Approach**

The analyses were conducted at the facet level and the whole-module level, respectively. The descriptive statistics were calculated using SPSS for windows V13.0. For hypothesis testing, a two-tailed probability value below .05 was considered statistically significant. The regression analysis (REG) was conducted on the whole-module level to understand which of the items were related to the dependent variable, where the dependent variables included the score of the first general item of WHOQOL-BREF, Q1 (“How would you rate your quality of life?”), the summed score of the first general items Q1 and the second general items Q2 of WHOQOL-BREF, Q1 + Q2 (Q2: “How satisfied are you with your health?”), and BREF total (the total score of WHOQOL-BREF), respectively, where we assumed the scores of Q1/Q2 and the total score of WHOQOL-BREF as the gold standards of QOL for the subject.

A principal components analysis (PCA) was also conducted on the whole-module level to reduce the number of variables and to detect the latent structure among variables, where an item was flagged for potential exclusion when it showed small loading, suggesting less importance for measurement of QOL. The corrected item–total correlation (Corr.r) was defined as the correlation between an item and the sum of other three items of its own facet, and it is automatically calculated within the SPSS reliability procedure. Confirmatory factor analysis (CFA) was used to assess how well the data fit the theoretical model, which was conducted using Mplus software (version 5.21; Muthén & Muthén, 1998–2007). Values of the comparative fit index (CFI) and non-normed fit index (NNFI) above 0.90 were
considered to represent an adequate fit, and 0.95 was generally accepted as reflecting good fit; the values of root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) less than 0.08 indicated a good fit (Hu & Bentler, 1999). Items were flagged if they had a low loading, which reflected the importance of an item to its own facet.

Item response theory (IRT) is a mathematical model-based approach used to understand the relationships between individuals’ parameters and their response patterns (Embretson & Reise, 2000). Compared with so-called classical test theory (CTT), IRT has many advantages such as sample independence, the estimates of parameters are measured on same logit scale, and it provides reliability at different levels of the underlying trait, item, or scale (Embretson & Reise, 2000). Based on the advantages mentioned earlier, IRT has been used to evaluate the psychometric properties of an existing scale, to optimally shorten a scale when necessary, and to evaluate the subsequent performance of the reduced scale. Usually in health outcomes research, the item responses are polytomous and ordered, so either the partial credit model (PCM; Masters, 1982) or the graded response model (GRM; Samejima, 1997) is the suitable model to use. In this study, both models were analyzed using different statistical software. The WHOQOL-OLD data were fitted to the PCM using the Rasch unidimensional measurement model software 2020 (RUMM; Andrich, Sheridan, Lyne, & Luo, 2003) and WINSTEPS. In the RUMM analysis, the item–trait interaction chi-square fit statistic was used to evaluate the adequacy of the fit of the facets to the Rasch model, and the fit of the items were also assessed through the individual item chi-square fit statistics. A significant $p$ value corresponding to a chi-square fit statistic below .01 was taken to indicate significant misfit, whereas a $p$ value between .01 and .05 was taken to indicate borderline misfit. In addition, RUMM analysis also provides a fit residual for each item. A residual beyond $\pm 2.5$ is considered to exhibit an item misfit to the Rasch model. In our study, a rule with a $p$ value below .01 and the absolute value of fit residual at a threshold of 2.5 was used to evaluate the item–model fitness. When all $p$ values were better than .01, the chi-square/degrees of freedom ($df$) values were also used to screen items such that the lower the ratio the better the item fit.

In WINSTEPS analysis, the mean squares (MnSq) were used to determine the extent to which the items contribute to the measurement model. The infit MnSq is an inliers pattern-sensitive fit statistic, which is more sensitive to unexpected patterns of observations, whereas the outfit MnSq is an outlier-sensitive fit statistic, which is more sensitive to unexpected values of observations. The ideal MnSq value is 1.0, reflecting no unexpected variance in the responses to the item. In this study, we considered an item with MnSq value beyond the interval (0.7, 1.3) as misfit (Wright, Linacre, Gustafson, & Martin-Lof, 1994), which might be either redundant or adding noise to the latent trait. DIF occurs when people from different groups with the same latent trait have a different probability of giving a certain response on a scale (Embretson & Reise, 2000). In our study, DIF analysis was performed across groups of gender (male vs. female), age ($\leq 80$ vs. $>80$ years), health (health vs. non-health), and center (European vs. non-European) using WINSTEPS software. By WINSTEPS, the item–difficulty measures were estimated at each group by anchoring person–ability measures to that of the overall sample, thereby ensuring that items were calibrated to the same scale across different populations. A criterion of .5 logits between item difficulties across groups was applied to determine whether an item exhibited DIF (Kendel et al., 2010).

In MULTILOG analysis, the values of discrimination parameter ($\alpha$) and item information function were estimated for each 5-category item. The discrimination parameter is the ability of an item to discriminate people with different levels of the underlying trait, and the total information was calculated by summing up the item information values, of which the greater the value the more contribution to the measure of facet.

In order to produce a best six-item short form of the OLD module, with one item per facet, the following three approaches were tested.

**First Approach (IRT and REG)**

The first approach included REG, PCA, and corrected item–total correlation analysis followed by Rasch analysis. In the REG within any facet of four items, the ranks of 1–4 refer to the items with the highest standardized regression coefficient
(Rank 1) to the lowest one (Rank 4), respectively. When REG and PCA were conducted at the whole-module level, the standardized regression coefficients and factor loadings were ranked from the highest value (Rank 1) to lowest value (Rank 24), respectively. In addition, the corrected item–total correlation coefficients within any facet were ranked from the highest value (Rank 1) to lowest value (Rank 4). After summing up the ranks of the earlier analyses for each item, about 10–12 items (at least one item per facet) with a lower rank sum were selected to be taken forward for Rasch analysis. Missing values were replaced with appropriate median scores prior to the Rasch analysis, so as to retain the 5-point response category structure for items. Then six items were retained according to the fit residual from several iterations of Rasch analysis. The short form resulted from the first approach was called the version 1 short-form WHOQOL-OLD module.

Second Approach (CTT and REG)

The second approach included REG within each facet and CFA and corrected item–total correlation analysis at the whole-module level. The rank of each REG was the same as the first approach mentioned earlier. When several REGs were used simultaneously, the REG ranks of each item were summed up to generate an REG rank sum as a summarized indicator of that item based on REG. In CFA, the loadings of the four items within each facet were ranked from the highest value (Rank 1) to the lowest value (Rank 4) and defined as the CFA ranks of the four items, respectively. The ranks of corrected item–total correlation coefficients within any facet were the same as mentioned earlier. Then a CTT rank sum was calculated through summing up of the REG rank sum, the CFA rank, and the rank of corrected item–total correlation coefficients for each item. Finally, the version 2 short-form WHOQOL-OLD module was formed through selecting the items with the lowest (CTT + IRT) rank sum within their own facets.

Third Approach (CTT and IRT and REG)

The third approach included the earlier mentioned calculation of CTT rank sum together with three IRT analyses carried out by RUMM, WINSTEPS, and MULTILOG. In the RUMM analysis, the fit residual and ChiSq/df values were ranked for the four items in each facet from the lowest absolute value (Rank 1) to the highest absolute value (Rank 4), respectively; then the two ranks of the fit residual and ChiSq/df values were summed to produce an RUMM rank sum for each item. In the WINSTEPS analysis, the absolute values of (1 − MnSq) of four items within any facet were ranked from the lowest (Rank 1) to the highest (Rank 4) for infit and outfit, respectively, and then the ranks for infit and outfit of each item were summed up to produce a WINSTEPS rank sum for the item. In the MULTILOG analysis, the values of four items in each facet on the discrimination parameter and item total information were ranked from the highest (Rank 1) to the lowest (Rank 4), respectively, and then the two ranks on the discrimination parameter and item total information were summed up to produce a MULTILOG rank sum. Then, the three rank sums (the RUMM rank sum, WINSTEPS rank sum, and MULTILOG rank sum) were summed up to produce an IRT rank sum. Finally, the CTT rank sum mentioned earlier by the second approach and the IRT rank sum here were summed up into a (CTT + IRT) rank sum, and the version 3 short-form WHOQOL-OLD module was formed through selecting the items with the lowest (CTT + IRT) rank sum within their own facets.

Validation of Three Versions of the Short-Form WHOQOL-OLD Module

We validated the three versions of the short-form scale using the validation sample by the following procedures: first, the correlations between the total scores of the short-form scale and WHOQOL-BREF and SF-12 scales were calculated; second, the internal consistency reliability values of the short-form scales were calculated; third, the multiple linear REGs were conducted to assess criterion validity using the total scores of WHOQOL-OLD and WHOQOL-BREF as dependent variables, respectively; fourth, the discriminant validity analyses were run to determine whether the short-form scales could discriminate the different levels of old adults; fifth, DIF analyses were conducted to examine the equivalence of item function; and sixth, the RUMM analyses (Rasch analyses) of the three versions were conducted to examine the unidimensionality.

Results

The descriptive information in Table 1 was generated for the developmental sample (N = 3,711)
and the validation sample (N = 1,855). The sociodemographic characteristics, the means, and standard deviations of the total scores of WHOQOL-OLD, WHOQOL-BREF, and SF-12 for the developmental and validation samples were similar. On the facet level, the differences between two samples were not significant at the .05 level (two-tailed tests) except for the “social participation” facet of WHOQOL-OLD.

Before the REGs based on facet level or whole-module level, to check the colinearity among items, the variance inflation factors (VIF) of all 24 items were calculated, respectively. As a result, the maximal VIF was 2.562, much less than 4. In addition, the Spearman’s correlation coefficients of four items within each facet were also calculated, of which their values were mostly within the interval (.3, .6). All these showed that the problem of colinearity in the multiple REGs could be ignored.

Table 2 lists the results from REG, PCA, and corrected item–total correlation analysis. In the REG on facet level, when Q1 (first general item on WHOQOL-BREF) was the dependent variable, OLD_01, OLD_09, and OLD_23 did not enter into the regression equation, whereas OLD_20, OLD_11, OLD_12, OLD_19, OLD_17, and OLD_21 showed greater contributions to their own facets; when Q1 + Q2 (summing up the first and second general items) was the dependent variable, nine items (OLD_01, OLD_10, OLD_03, OLD_05, OLD_06, OLD_09, OLD_16, OLD_21, and OLD_23) did not enter into the equation; when Q1 + Q2 was the dependent variable, nine items...

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sample (N)</th>
<th>Statistics t/she ( \chi^2 )</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Developmental (3,711)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤80</td>
<td>2,902 (78.2)</td>
<td>1.73</td>
<td>.19</td>
</tr>
<tr>
<td>&gt;80</td>
<td>809 (21.8)</td>
<td>0.37</td>
<td>.55</td>
</tr>
<tr>
<td>Gender, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,543 (41.8)</td>
<td>0.32</td>
<td>.57</td>
</tr>
<tr>
<td>Female</td>
<td>2,147 (58.2)</td>
<td>1.03</td>
<td>.32</td>
</tr>
<tr>
<td>Healthy, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td>2,547 (71.2)</td>
<td>2.67</td>
<td>.10</td>
</tr>
<tr>
<td>Unhealthy</td>
<td>1,029 (28.8)</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Center, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>2,263 (61.0)</td>
<td>1.69</td>
<td>.09</td>
</tr>
<tr>
<td>Non-European</td>
<td>1,448 (39.0)</td>
<td>2.56</td>
<td>.05</td>
</tr>
<tr>
<td>WHOQOL-OLD, Mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>66.5 (13.7)</td>
<td>65.7 (13.6)</td>
<td>1.69</td>
</tr>
<tr>
<td>Sensory abilities (OLD-SAB)</td>
<td>72.8 (21.6)</td>
<td>72.1 (21.7)</td>
<td>1.12</td>
</tr>
<tr>
<td>Autonomy (OLD-AUT)</td>
<td>66.7 (17.9)</td>
<td>66.0 (17.8)</td>
<td>1.35</td>
</tr>
<tr>
<td>Death and dying (OLD-DAD)</td>
<td>63.9 (24.8)</td>
<td>63.1 (25.1)</td>
<td>1.13</td>
</tr>
<tr>
<td>Past, present, and future activities (OLD-PPF)</td>
<td>64.5 (17.4)</td>
<td>63.8 (17.3)</td>
<td>1.39</td>
</tr>
<tr>
<td>Social participation (OLD-SOP)</td>
<td>66.5 (18.4)</td>
<td>65.3 (18.5)</td>
<td>2.28</td>
</tr>
<tr>
<td>Intimacy (OLD-INT)</td>
<td>62.3 (24.4)</td>
<td>61.8 (24.6)</td>
<td>0.78</td>
</tr>
<tr>
<td>WHOQOL-BREF, Mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>67.3 (14.3)</td>
<td>66.7 (14.2)</td>
<td>1.50</td>
</tr>
<tr>
<td>Physical</td>
<td>66.6 (19.0)</td>
<td>66.2 (19.1)</td>
<td>0.74</td>
</tr>
<tr>
<td>Psychological</td>
<td>66.7 (15.7)</td>
<td>66.0 (15.7)</td>
<td>1.50</td>
</tr>
<tr>
<td>Social</td>
<td>65.9 (17.4)</td>
<td>65.3 (17.4)</td>
<td>1.21</td>
</tr>
<tr>
<td>Environment</td>
<td>69.0 (16.0)</td>
<td>68.2 (16.0)</td>
<td>1.78</td>
</tr>
<tr>
<td>SF-12, Mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>62.2 (15.9)</td>
<td>61.4 (16.3)</td>
<td>1.32</td>
</tr>
<tr>
<td>Physical component summary</td>
<td>65.2 (24.6)</td>
<td>64.0 (25.4)</td>
<td>1.29</td>
</tr>
<tr>
<td>Mental component summary</td>
<td>59.9 (13.4)</td>
<td>59.3 (13.5)</td>
<td>1.32</td>
</tr>
</tbody>
</table>

Notes: *Chi-square test for age, gender, healthy, and center; t test for others. The bold value .02 showed statistical significance at the test level of .05.
Table 2. Results of Regression Analysis (REG), Correlation Analysis, and PCA, Including Standardized Regression Coefficients, Correlation Coefficients, Factor Loadings, and Corresponding Ranks, Resulting in the Items Forward to the Last Step of the First Approach: Further Rasch Analysis (N = 3,711)

<table>
<thead>
<tr>
<th>Item/facet</th>
<th>Sensory abilities (OLD-SAB)</th>
<th>Autonomy (OLD-AUT)</th>
<th>Death and dying (OLD-DAD)</th>
<th>Past, present, and future activities (OLD-PPF)</th>
<th>Social participation (OLD-SOP)</th>
<th>Intimacy (OLD-INT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLD_01</td>
<td>— (4)</td>
<td>.06 (4)</td>
<td>.07 (2)</td>
<td>.11 (3)</td>
<td>.17 (2)</td>
<td>— (4)</td>
</tr>
<tr>
<td>OLD_02</td>
<td>— (4)</td>
<td>.18 (2)</td>
<td>.06 (3)</td>
<td>.09 (4)</td>
<td>.07 (3)</td>
<td>— (4)</td>
</tr>
<tr>
<td>OLD_10</td>
<td>.05 (3)</td>
<td>.04 (3)</td>
<td>.06 (3)</td>
<td>.06 (3)</td>
<td>.06 (3)</td>
<td>— (4)</td>
</tr>
<tr>
<td>OLD_20</td>
<td>.29 (1)</td>
<td>.33 (1)</td>
<td>.33 (1)</td>
<td>.40 (1)</td>
<td>.45 (4)</td>
<td>— (4)</td>
</tr>
<tr>
<td>OLD_03</td>
<td>.06 (4)</td>
<td>.11 (4)</td>
<td>.17 (1)</td>
<td>.41 (1)</td>
<td>.45 (4)</td>
<td>— (4)</td>
</tr>
<tr>
<td>OLD_04</td>
<td>.20 (2)</td>
<td>.21 (2)</td>
<td>.20 (2)</td>
<td>.09 (2)</td>
<td>.66 (3)</td>
<td>— (4)</td>
</tr>
<tr>
<td>OLD_05</td>
<td>.12 (3)</td>
<td>.11 (3)</td>
<td>.17 (3)</td>
<td>.06 (3)</td>
<td>.56 (4)</td>
<td>— (4)</td>
</tr>
<tr>
<td>OLD_11</td>
<td>.29 (1)</td>
<td>.36 (1)</td>
<td>.40 (1)</td>
<td>.40 (1)</td>
<td>.45 (4)</td>
<td>— (4)</td>
</tr>
<tr>
<td>OLD_06</td>
<td>.07 (2)</td>
<td>.09 (1)</td>
<td>.17 (1)</td>
<td>.74 (1)</td>
<td>— (20)</td>
<td>— (20)</td>
</tr>
<tr>
<td>OLD_07</td>
<td>.06 (3)</td>
<td>.06 (3)</td>
<td>— (4)</td>
<td>.73 (2)</td>
<td>.05 (12)</td>
<td>.08 (8)</td>
</tr>
<tr>
<td>OLD_08</td>
<td>.10 (1)</td>
<td>.07 (2)</td>
<td>.09 (2)</td>
<td>.66 (3)</td>
<td>.04 (14.5)</td>
<td>— (20)</td>
</tr>
<tr>
<td>OLD_09</td>
<td>— (4)</td>
<td>.05 (4)</td>
<td>.06 (3)</td>
<td>.56 (4)</td>
<td>— (20)</td>
<td>— (20)</td>
</tr>
<tr>
<td>OLD_12</td>
<td>.27 (2)</td>
<td>.30 (2)</td>
<td>.35 (1)</td>
<td>.55 (1)</td>
<td>.11 (2)</td>
<td>.12 (3.5)</td>
</tr>
<tr>
<td>OLD_13</td>
<td>.11 (3)</td>
<td>.08 (3)</td>
<td>.18 (3)</td>
<td>.53 (3)</td>
<td>.05 (12)</td>
<td>.20 (8)</td>
</tr>
<tr>
<td>OLD_15</td>
<td>.08 (4)</td>
<td>.05 (4)</td>
<td>.10 (4)</td>
<td>.50 (4)</td>
<td>.04 (14.5)</td>
<td>— (20)</td>
</tr>
<tr>
<td>OLD_19</td>
<td>.31 (1)</td>
<td>.33 (1)</td>
<td>.31 (1)</td>
<td>.54 (2)</td>
<td>.16 (1)</td>
<td>.15 (1)</td>
</tr>
<tr>
<td>OLD_14</td>
<td>.17 (3)</td>
<td>.12 (3)</td>
<td>.20 (3)</td>
<td>.52 (4)</td>
<td>.06 (9)</td>
<td>.05 (12.5)</td>
</tr>
<tr>
<td>OLD_16</td>
<td>.15 (4)</td>
<td>.16 (4)</td>
<td>.19 (4)</td>
<td>.65 (2)</td>
<td>— (20)</td>
<td>.06 (10)</td>
</tr>
<tr>
<td>OLD_17</td>
<td>.20 (1)</td>
<td>.26 (1)</td>
<td>.26 (1)</td>
<td>.70 (1)</td>
<td>.10 (3)</td>
<td>.12 (3.5)</td>
</tr>
<tr>
<td>OLD_18</td>
<td>.19 (2)</td>
<td>.19 (2)</td>
<td>.21 (2)</td>
<td>.58 (3)</td>
<td>.08 (4.5)</td>
<td>.08 (6.5)</td>
</tr>
<tr>
<td>OLD_21</td>
<td>.20 (1)</td>
<td>.22 (1)</td>
<td>.28 (1)</td>
<td>.66 (4)</td>
<td>— (20)</td>
<td>.05 (12.5)</td>
</tr>
<tr>
<td>OLD_22</td>
<td>.11 (3)</td>
<td>.10 (3)</td>
<td>.13 (2)</td>
<td>.78 (1)</td>
<td>.05 (12)</td>
<td>.04 (14.5)</td>
</tr>
<tr>
<td>OLD_23</td>
<td>— (4)</td>
<td>.05 (4)</td>
<td>.04 (2)</td>
<td>.76 (3)</td>
<td>— (20)</td>
<td>.04 (20)</td>
</tr>
<tr>
<td>OLD_24</td>
<td>.14 (2)</td>
<td>.11 (2)</td>
<td>.11 (3)</td>
<td>.77 (2)</td>
<td>.06 (9)</td>
<td>.20 (2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REG-Q1 (rank)</th>
<th>REG-Q1 + 2 (rank)</th>
<th>REG-BREF (rank)</th>
<th>Corr.r (rank)</th>
<th>REG-Q1 (rank)</th>
<th>REG-Q1 + 2 (rank)</th>
<th>REG-BREF (rank)</th>
<th>PCA (Rank)</th>
<th>V.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>.836&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.837&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.736&lt;sup&gt;a&lt;/sup&gt;</td>
<td>— (20)</td>
<td>— (20)</td>
<td>— (20)</td>
<td>— (20)</td>
<td>— (20)</td>
<td>— (20)</td>
</tr>
</tbody>
</table>

Notes: <sup>a</sup>The internal consistency reliability coefficients of six subscales. Regression equations are for dependent variables Q1 (first general item on the BREF), Q1 + Q2 (first and second general items summed), and BREF total score. The mark “—” means that variables did not enter regression equations or showed lower loadings. Numbers in columns show the rank of the item contributing most (=1) to least (=4) in a facet or in the overall scale. Corr.<sup>r</sup> = corrected item–total correlation coefficients; PCA = principal components analysis for the whole module with 24 items.

OLD_01, OLD_10, OLD_05, OLD_08, OLD_09, OLD_13, OLD_15, OLD_23, and OLD_24 did not enter into the equation; and when the BREF total score was the dependent variable, seven items (OLD_10, OLD_04, OLD_08, OLD_09, OLD_14, OLD_23, and OLD_24) did not enter into the equation. As seen from the REGs mentioned earlier, items OLD_01, OLD_10, OLD_09, and OLD_23 displayed little or no contribution to the measure of QOL.

In PCA, six variables with a low loading were excluded. Except for items OLD_05 and OLD_11, the other corrected correlation coefficients were greater than .50. Cronbach’s <sup>a</sup> of six facets ranged from .724 to .876. The CFA model consisting of six facets with four items each showed good fit indices (χ<sup>2</sup> = 2,233.6, df = 237, CFI = 0.936; NNFI = 0.925, RMSEA = 0.051, SRMR = 0.043), and the standardized factor loadings for all items were positive and significant, ranging from 0.59 to 0.93.

**IRT Analysis**

The utility of the IRT model is dependent on the extent to which the model accurately reflects the data or the degree of model data fit. All applications of IRT need to satisfy the two assumptions of
unidimensionality and local independence. To test these assumptions, EFA and CFA were conducted. The results showed that the ratios of the eigenvalues of the first and second factors all were greater than 3, which indicated that they measured the unique trait of related facets, and the maximum correlation for the residuals was .084, that is, all correlations were less than .20. This finding supported the use of IRT (Reeve et al., 2007; Slocum, 2005). For each item, all the differences on the parameter of “difficulty” across the earlier contrast groups of gender, age, health, and center fell below the criterion of .5 logit so that not any item with significant DIF was found.

Table 3 displays the detailed results for the IRT analyses together with their ranks.

In the RUMM analysis, only eight items (OLD_04, OLD_05, OLD_08, OLD_12, OLD_13, OLD_15, OLD_19, and OLD_18) showed good fit to the model with their fit residuals within the interval (−2.5, 2.5).

In the WINSTEPS analysis, except item OLD_10 of sensory abilities facet, item OLD_09 of death and dying facet, and item OLD_21 of intimacy facet (with values 1.33, 1.33, and 1.31, respectively), the infit statistics of the remaining items fell within the specified range (0.7, 1.3). Except item OLD_02 of sensory abilities facet, item OLD_14 of social participation facet, and item OLD_21 of intimacy facet (with values 0.65, 1.32, and 1.32, respectively), the outfit statistics of the remaining items fell in the interval (0.7, 1.3). In general, items OLD_02, OLD_10, OLD_09, OLD_14, and OLD_21 did not fit a single construct, and the remaining items all satisfied the unidimensionality assumption of the IRT model.

In the MULTILOG analysis, expect OLD_11 with a moderate discrimination value (<1.34), the remaining items showed high discrimination, which ranged from 1.35 to 4.63. There were four items (OLD_11, OLD_09, OLD_15, and OLD_14) with the amount of information below 20, which suggested lower effects for their own facets.

Version 1 —From First Approach

On the basis of the ranks of REG, PCA, and correlation analyses in the Table 2, 10 items (items OLD_04, OLD_06, OLD_11, OLD_12, OLD_16, OLD_17, OLD_19, OLD_20, OLD_21, and OLD_24) were selected to be taken forward to Rasch analysis according to lower ranks and selection in PCA. The poorest fit on several indicators was OLD_06, but it was not clear that any other item from the death and dying facet would perform better. Therefore, OLD_06 was retained despite some reservations. Due to items OLD_21 and OLD_24 belonging to the same facet, item OLD_24, which showed poorer fit, was excluded. The remaining nine items were examined with three iterations of Rasch analysis. OLD_04 showed reversed thresholds, suggesting unreasonable options, and OLD_17 and OLD_19 showed poorer fit so that these three items also were deleted. Finally, the six items, OLD_06, OLD_11, OLD_12, OLD_16, OLD_20, and OLD_21, formed the version 1 short-form WHOQOL-OLD, which demonstrated a relative good model fit (chi-squared \(df\) = 854.8 [54]; \(p < .001\)).

Version 2 —From Second Approach

In order to avoid a possible bias for the REG, the three ranks of REGs at the facet level listed in Table 3 were summed up into an REG sum. The REG sum, ranks of corrected item–total correlations, and ranks of factor loadings were summed up into a CTT rank sum, as shown in Table 4. According to the CTT rank sum, seven items with the lowest rank in each facet were retained, which included OLD_02, OLD_04, OLD_11, OLD_06, OLD_12, OLD_17, and OLD_22. Due to items OLD_04 and OLD_11, both being in the “autonomy” facet, two short forms including OLD_04 or OLD_11 were compared by Cronbach’s \(\alpha\) coefficients. As a result, the one including OLD_11 showed better internal reliability than that including OLD_04 (.669 vs. .644). Therefore, the version 2 short-form WHOQOL-OLD consisted of the following items: OLD_02, OLD_11, OLD_06, OLD_12, OLD_17, and OLD_22.

Version 3 —From Third Approach

When the IRT rank sum was considered with the CTT rank sum together (i.e., the two ranks summed up), seven items (OLD_20, OLD_04, OLD_06, OLD_19, OLD_17, OLD_18, and OLD_24) with the lowest rank in each facet were potentially selected, which have been listed in the last column of Table 4. Due to items OLD_17 and OLD_18, both being in the social participation facet, a short form including OLD_17 and another form including OLD_18 were formed. The former displayed better internal reliability than the latter, with Cronbach’s \(\alpha\) coefficients .646 and .637,
Table 3. Detailed Results of the WHOQOL-OLD Scale Analysis Based on IRT and the Rank of Items According to Fit Indices (N = 3,711)

<table>
<thead>
<tr>
<th>Facet/item</th>
<th>RUMM Fit residual (rank)</th>
<th>ChiSq(df/rank)</th>
<th>RUMM Rank sum 1 (1)</th>
<th>WINSTEPS InfitmNsq (rank)</th>
<th>Outfit MnSq (rank)</th>
<th>WINSTEPS Rank sum 2 (2)</th>
<th>MULTILOG Discrimination (rank)</th>
<th>Item total information (rank)</th>
<th>MULTILOG Rank sum 3 (3)</th>
<th>IRT rank sum (1) + (2) + (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensory abilities (OLD-SAB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_01</td>
<td>−4.88 (3)</td>
<td>26.66 (4)</td>
<td>7</td>
<td>0.82 (2)</td>
<td>0.78 (3)</td>
<td>5</td>
<td>3.79 (2)</td>
<td>65.12 (2)</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>OLD_02</td>
<td>−8.43 (4)</td>
<td>23.85 (3)</td>
<td>7</td>
<td>0.71 (3)</td>
<td>0.65 (4)</td>
<td>7</td>
<td>4.63 (1)</td>
<td>80.53 (1)</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>OLD_10</td>
<td>3.05 (2)</td>
<td>21.23 (2)</td>
<td>4</td>
<td>1.33 (4)</td>
<td>1.18 (2)</td>
<td>6</td>
<td>1.89 (4)</td>
<td>20.62 (4)</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>OLD_20</td>
<td>2.54 (1)</td>
<td>15.44 (1)</td>
<td>2</td>
<td>1.14 (1)</td>
<td>1.13 (1)</td>
<td>2</td>
<td>2.01 (3)</td>
<td>28.17 (3)</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Autonomy (OLD-AUT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_03</td>
<td>−4.66 (4)</td>
<td>11.02 (4)</td>
<td>8</td>
<td>0.87 (4)</td>
<td>0.83 (4)</td>
<td>8</td>
<td>2.31 (1)</td>
<td>31.89 (1)</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>OLD_04</td>
<td>−0.37 (2)</td>
<td>4.61 (2)</td>
<td>4</td>
<td>0.97 (2)</td>
<td>0.97 (2)</td>
<td>4</td>
<td>1.81 (2.5)</td>
<td>24.40 (2)</td>
<td>4.5</td>
<td>12.5</td>
</tr>
<tr>
<td>OLD_05</td>
<td>0.04 (1)</td>
<td>3.22 (1)</td>
<td>2</td>
<td>1.01 (1)</td>
<td>1.00 (1)</td>
<td>2</td>
<td>1.81 (2.5)</td>
<td>23.24 (3)</td>
<td>5.5</td>
<td>9.5</td>
</tr>
<tr>
<td>OLD_11</td>
<td>4.53 (3)</td>
<td>6.17 (3)</td>
<td>6</td>
<td>1.13 (3)</td>
<td>1.15 (3)</td>
<td>6</td>
<td>1.28 (4)</td>
<td>13.45 (4)</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>Death and dying (OLD-DAD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_06</td>
<td>−3.82 (2)</td>
<td>13.07 (4)</td>
<td>6</td>
<td>0.82 (3)</td>
<td>0.81 (2)</td>
<td>5</td>
<td>3.56 (1)</td>
<td>57.96 (1)</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>OLD_07</td>
<td>−4.40 (3)</td>
<td>11.61 (3)</td>
<td>6</td>
<td>0.83 (2)</td>
<td>0.80 (3)</td>
<td>5</td>
<td>3.22 (2)</td>
<td>48.44 (2)</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>OLD_08</td>
<td>0.25 (1)</td>
<td>4.47 (1)</td>
<td>2</td>
<td>1.02 (1)</td>
<td>1.01 (1)</td>
<td>2</td>
<td>2.35 (3)</td>
<td>31.00 (3)</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>OLD_09</td>
<td>6.95 (4)</td>
<td>6.70 (2)</td>
<td>6</td>
<td>1.33 (4)</td>
<td>1.30 (4)</td>
<td>8</td>
<td>1.59 (4)</td>
<td>19.07 (4)</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>Past, present, and future activities (OLD-PPF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_12</td>
<td>−1.49 (4)</td>
<td>1.88 (3)</td>
<td>7</td>
<td>0.98 (1)</td>
<td>0.96 (4)</td>
<td>5</td>
<td>1.82 (2)</td>
<td>24.59 (2)</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>OLD_13</td>
<td>−0.90 (2)</td>
<td>3.06 (4)</td>
<td>6</td>
<td>0.97 (2.5)</td>
<td>0.97 (2)</td>
<td>4.5</td>
<td>1.77 (3)</td>
<td>23.73 (3)</td>
<td>6</td>
<td>16.5</td>
</tr>
<tr>
<td>OLD_15</td>
<td>0.99 (3)</td>
<td>1.88 (2)</td>
<td>5</td>
<td>1.05 (4)</td>
<td>1.03 (2)</td>
<td>6</td>
<td>1.59 (4)</td>
<td>18.32 (4)</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>OLD_19</td>
<td>−0.65 (1)</td>
<td>1.80 (1)</td>
<td>2</td>
<td>0.97 (2.5)</td>
<td>0.97 (2)</td>
<td>4.5</td>
<td>1.86 (1)</td>
<td>25.10 (1)</td>
<td>2</td>
<td>8.5</td>
</tr>
<tr>
<td>Social participation (OLD-SOP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_14</td>
<td>6.37 (3)</td>
<td>11.23 (3)</td>
<td>6</td>
<td>1.22 (3)</td>
<td>1.32 (4)</td>
<td>7</td>
<td>1.35 (4)</td>
<td>13.72 (4)</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>OLD_16</td>
<td>−3.32 (2)</td>
<td>8.06 (2)</td>
<td>4</td>
<td>0.89 (2)</td>
<td>0.87 (2)</td>
<td>4</td>
<td>2.68 (2)</td>
<td>42.54 (2)</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>OLD_17</td>
<td>−7.62 (4)</td>
<td>14.33 (4)</td>
<td>8</td>
<td>0.76 (4)</td>
<td>0.74 (3)</td>
<td>7</td>
<td>4.03 (1)</td>
<td>74.26 (1)</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>OLD_18</td>
<td>1.78 (1)</td>
<td>1.64 (1)</td>
<td>2</td>
<td>1.08 (1)</td>
<td>1.08 (1)</td>
<td>2</td>
<td>1.82 (3)</td>
<td>25.01 (3)</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Intimacy (OLD-INT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_21</td>
<td>7.03 (4)</td>
<td>7.28 (4)</td>
<td>8</td>
<td>1.31 (4)</td>
<td>1.32 (4)</td>
<td>8</td>
<td>2.00 (4)</td>
<td>26.97 (4)</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>OLD_22</td>
<td>−3.55 (2)</td>
<td>5.65 (2)</td>
<td>4</td>
<td>0.87 (2.5)</td>
<td>0.87 (2)</td>
<td>4.5</td>
<td>3.29 (3)</td>
<td>51.45 (3)</td>
<td>6</td>
<td>14.5</td>
</tr>
<tr>
<td>OLD_23</td>
<td>−4.01 (3)</td>
<td>5.74 (3)</td>
<td>6</td>
<td>0.90 (1)</td>
<td>0.87 (2)</td>
<td>3</td>
<td>3.72 (1)</td>
<td>57.64 (2)</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>OLD_24</td>
<td>−3.26 (1)</td>
<td>4.93 (1)</td>
<td>2</td>
<td>0.87 (2.5)</td>
<td>0.87 (2)</td>
<td>4.5</td>
<td>3.58 (2)</td>
<td>58.28 (1)</td>
<td>3</td>
<td>9.5</td>
</tr>
</tbody>
</table>

Notes: RUMM and WINSTEPS analyses were based on partial credit model, and MULTILOG analysis was based on graded response model. ChiSq(df) = the ratio of the fit statistic and df; IRT = item response theory; rank sum 1 = the sum of the ranks for fit residual and ChiSq(df); rank sum 2 = the sum of the ranks for |1 − infit MnSq| and |1 − outfit MnSq|; rank sum 3 = the sum of the ranks for discrimination and item total information; RUMM = Rasch unidimensional measurement model.
Table 4. Ranks of the Items Within Their Facet Resulted From the Methods Concerned (N = 3,711)

<table>
<thead>
<tr>
<th>Item/facet</th>
<th>REG</th>
<th>CTT and REG</th>
<th>CTT and IRT and REG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensory abilities (OLD-SAB)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_01</td>
<td>4</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>OLD_02</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>OLD_10</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>OLD_20</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Autonomy (OLD-AUT)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_03</td>
<td>4</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>OLD_04</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>OLD_05</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>OLD_11</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Death and dying (OLD-DAD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_06</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>OLD_07</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>OLD_09</td>
<td>4</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Past, present, and future activities (OLD-PPF)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_12</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>OLD_13</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>OLD_15</td>
<td>4</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>OLD_19</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Social participation (OLD-SOP)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_14</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>OLD_16</td>
<td>4</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>OLD_17</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>OLD_18</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Intimacy (OLD-INT)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD_21</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>OLD_22</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>OLD_23</td>
<td>4</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>OLD_24</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Notes: Regression equations at facet level are for dependent variables Q1 (first general item on BREF), Q1 + Q2 (first and second general items summed), and BREF total score. Numbers in columns show rank of item contributing most (=1) to least variance (=4) in facet. IRT rank sum comes from the last column in Table 3. CFA = confirmatory factor analysis; CTT = classical test theory; IRT = item response theory; REG, regression analysis.
respectively. Therefore, the former was selected as the version 3 short form of OLD module, which included the following items: OLD_20, OLD_04, OLD_06, OLD_19, OLD_17, and OLD_24.

Validation of the Three Versions of the Short Form

The detailed item content of the three versions of the short forms is shown in Table 5. The correlations between the total scores of the three versions and those of WHOQOL-OLD, WHOQOL-BREF, and SF-12, and the reliability values based on the validation sample are shown in Table 6. The reliability values of the three versions were all greater than .6, indicating acceptable reliability. Compared with version 2 or version 3, version 1 showed the best internal reliability, with Cronbach’s $\alpha = .681$, with version 2 next.

In the correlation analyses, version 2 had the highest correlation with the total score of the original WHOQOL-OLD module and version 3 next; version 1 had the highest correlation with the total score of WHOQOL-BREF and version 2 next; version 2 had the highest correlation with the total score of SF-12 and version 1 next. The three versions showed different degrees of correlation with the domains of the WHOQOL-BREF and SF-12. All correlation coefficients ranged from .45 to .80, suggesting the three versions had good concurrent validity. Version 1 showed high correlation with “Psychological” and “physical component summary” domains, and version 2 had high correlations with the “physical” and “physical component summary” domains; version 3 showed high correlations with the “social” and “environment” domains. To assess criterion validity, multiple linear REGs were conducted using the total scores of the original WHOQOL-OLD and WHOQOL-BREF as dependent variables, respectively. The results showed that all six items from any one of the three short versions were significant in the proposed models, and version 2 explained 86.4% of the variance of WHOQOL-OLD and 58.6% of the variance of WHOQOL-BREF. Version 3 explained 85.2% of the variance of WHOQOL-OLD and 58.8% of the variance of WHOQOL-BREF, and version 1 explained least percentage of

<table>
<thead>
<tr>
<th>WHOQOL-OLD</th>
<th>Short form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensory abilities (OLD-SAB)</td>
<td>OLD-01: Impairments to senses affect daily life</td>
</tr>
<tr>
<td></td>
<td>OLD-02: Loss of sensory abilities affect participation in activities</td>
</tr>
<tr>
<td></td>
<td>OLD-10: Problems with sensory functioning affect ability to interact</td>
</tr>
<tr>
<td></td>
<td>OLD-20: Rate sensory functioning</td>
</tr>
<tr>
<td>Autonomy (OLD-AUT)</td>
<td>OLD-03: Freedom to make own decisions</td>
</tr>
<tr>
<td></td>
<td>OLD-04: Feel in control of your future</td>
</tr>
<tr>
<td></td>
<td>OLD-05: People around you are respectful of your freedom</td>
</tr>
<tr>
<td></td>
<td>OLD-11: Able to do things you’d like</td>
</tr>
<tr>
<td></td>
<td>OLD-20: Rate sensory functioning</td>
</tr>
<tr>
<td>Death and dying (OLD-DAD)</td>
<td>OLD-06: Concerned about the way you will die</td>
</tr>
<tr>
<td></td>
<td>OLD-07: Afraid of not being able to control death</td>
</tr>
<tr>
<td></td>
<td>OLD-08: Scared of dying</td>
</tr>
<tr>
<td></td>
<td>OLD-09: Fear pain before death</td>
</tr>
<tr>
<td>Past, present, and future activities (OLD-PPF)</td>
<td>OLD-12: Satisfied with opportunities to continue achieving</td>
</tr>
<tr>
<td></td>
<td>OLD-13: Received the recognition you deserve in life</td>
</tr>
<tr>
<td></td>
<td>OLD-15: Satisfied with what you’ve achieved in life</td>
</tr>
<tr>
<td></td>
<td>OLD-19: Happy with things to look forward to</td>
</tr>
<tr>
<td>Social participation (OLD-SOP)</td>
<td>OLD-14: Have enough to do each day</td>
</tr>
<tr>
<td></td>
<td>OLD-16: Satisfied with the way you use your time</td>
</tr>
<tr>
<td></td>
<td>OLD-17: Satisfied with level of activity</td>
</tr>
<tr>
<td></td>
<td>OLD-18: Satisfied with opportunity to participate in community</td>
</tr>
<tr>
<td>Intimacy (OLD-INT)</td>
<td>OLD-21: Feel a sense of companionship in life</td>
</tr>
<tr>
<td></td>
<td>OLD-22: Experience love in your life</td>
</tr>
<tr>
<td></td>
<td>OLD-23: Opportunities to love</td>
</tr>
<tr>
<td></td>
<td>OLD-24: Opportunities to be loved</td>
</tr>
</tbody>
</table>

Notes: V.1, V.2, and V.3 mean the three versions of the short-form WHOQOL-OLD module; “√” indicates the item selected from the WHOQOL-OLD module to form the short form.
the variances of WHOQOL-OLD and WHOQOL-BREF. All correlation coefficients ranged from .45 to .80 and $R^2$ greater than .58 suggesting that the criterion validity of the three versions was acceptable.

To check the discriminant validity, the total scores of the three short-form versions were significantly different between male and female, between $\leq 80$ and $>80$ years, or between health and non-health old adults. However, the total score of version 1 was not significantly different between European and non-European old adults (see Table 6).

DIF analyses showed that OLD_20 of the short-form V3 had significant DIF with the difference on the parameter of difficulty across age group ($\leq 80$ vs. $>80$ years) equal to −.56, whereas the other items did not have significant DIF across gender, age group, health status, or center (see Table 6).

The RUMM analyses of the three versions were conducted to identify items with the best fit to a unidimensional model using the validation sample data set. Although the three versions showed item–trait interaction misfit ($p < .001$), version 3 showed relatively the best fit to the model, with the lowest chi-square statistics and fewest misfit items.

**Discussion**

The purpose of this article was to select six items from the initial WHOQOL-OLD module, which has 24 items, using CTT and then advanced IRT. Three possible short-form versions were created through three different analytic approaches.

In the first approach, several multiple linear regression models were used to determine the candidate items, which could explain more variance of the dependent variables, Q1, Q1 + Q2, and WHOQOL-BREF total score; then a simple Rasch analysis was used head-to-head comparing the candidate items in terms of IRT. The first approach was simpler, easier to operate, and utilized the advantage of IRT after using REG.

Because Q1, Q1 + Q2, and WHOQOL-BREF were correlated with each other, to avoid overweighting the results of REGs, in the second approach, the ranks from the REGs were integrated into a single rank sum (column 4 of Table 4). And
for similar reasons, the ranks by CTT were also integrated into one rank sum (column 7 of Table 4).

IRT is a very useful tool in constructing an efficient short-form scale. Its use enables the examination of item–performance and the identification and removal of poor items to improve the efficiency of the scale. The satisfaction of the requirements of unidimensionality and local independence can guarantee that the scale fits the expected model and the performance of the scale is stable and not dependent on the study sample. Our data showed good evidence supporting unidimensionality and model fit. In the IRT analyses in this paper, two models (the PCM and the GRM) with different properties and theoretical frameworks used simultaneously provided a more persuasive approach than only using one model and which thereby avoided the one-sidedness of items selected due to different analysis models. In the analyses based on PCM, due to the fact that the chi-square statistic and fit residual from RUMM analyses can be affected by the sample size (e.g., the chi-square statistics can be large even if the data fit to the model; Bollen, 1990), the infit and outfit indices from WINSTEPS analyses were also considered, which take expected variance and sample size into account. The use of two analyses simultaneously complements one another and may allow screening of more realistic and reliable items. In addition, one important criterion of item selection is “item information,” which helps quantify the process of item selection. The third approach was a more reasonable and widely used method through combining CTT and IRT methods to select the best items in each facet.

As shown in Table 6, the three versions showed similar reliability and validity, with the differences in reliability values, correlation coefficients, and $R^2$ among the three versions being minor, and any two versions shared two or three items in fact. Therefore, it was difficult to determine which version was the best. By now, we would recommend the three versions equally to the researchers who intend to apply the short forms of WHOQOL-OLD. In practice, the users could select one of the three versions according to the content of the items to meet their own specific needs. As we know, the version 1 has been used in a pilot study by three European countries so that its validation can be tested further.

There are a number of limitations with the analyses presented in this article. The initial WHOQOL-OLD module is a multidimensional scale. In using unidimensional analysis for each facet, one might lose some information about the correlation among the facets. The short-form scale was validated only using the original investigation data of the year 2005 because of the current availability of these data. The evaluation of the performance of the short-form versions was not based on a completely independent sample of respondents who were surveyed only with the short-form scale. Therefore, more studies on the validation of these three versions of the short-form WHOQOL-OLD module will be necessary with new data sets in order to allow implementation in future international studies.

Nevertheless, containing the relatively best items from the original scale, each of the three versions of the short-form WHOQOL-OLD covers six facets of traits about older adults’ health with six items only; they show acceptable internal consistency reliability and criterion validity; and they are more convenient in order to monitor the changes in health for older adults at any time and thereby supply the relevant information for doctors and for family members.

Funding

European Commission Fifth Framework (QLRT-2000-00320).

Acknowledgments

The study was carried out under the auspices of the World Health Organization Quality of Life Group (WHOQOL Group). The WHOQOL-OLD Group comprises a coordinating group and collaborating investigators in each of the following field centers: Professor M. Power, K. Quinn, K. Laidlaw, H. Toner, University of Edinburgh, UK; Dr R. Lucas, Institute Catala’ de l’Enveliment, Barcelona, Spain; Professor S. Skevington, Dr F. McCrate, University of Bath, UK; Professor Marianne Amir (we note with great sadness the death of Professor Marianne Amir in January 2004), Y. Ben Ya’acov, Tal Narkiss-Guez, Department of Behavioural Sciences, Ben-Gurion University of the Negev, Israel; Professor K. Lunnan, National institute for Medical Rehabilitation, Hungary; Dr K. Martigny, G. Bech-Anderson, Psychiatric Research Unit, Copenhagen, Denmark; Professor Ji-Qian Fang, Dr Yuantaao Hao, Sun Yat-sen University of Medical Sciences, Guangzhou, China; Professor M. C. Angermeyer, I. Winkler, Department of Psychiatry, University of Leipzig, Germany; Professor G. Hawthorne, Australian Centre for Posttraumatic Health, University of Melbourne, Australia; Dr M. Kalfoss, Faculty of Nursing, Menighetssoster-hjemmet College, Oslo, Norway; Dr A. Leplege, INSERM, Paris, France; Dr E. Dragomirecka, Prague Psychiatric Centre, Prague, Czech Republic; Dr M. Martin and Mr. D. Bushnell, Health Services Research Associates Inc., Seattle, USA; Dr M. Tazaki, Department of Science, Science University of Tokyo, Tokyo, Japan; Professor M. Eiseman, Department of Psychology, University of Tromsoe, Norway; B. Nygren, Department of Nursing, Umea University, Sweden; Dr A. Molfzan, Faculty of Human and Social Development, University of Victoria, Canada; Dr J. Ceremynych, Scientific Department of Gerontology Problems, Institute of Experimental and Clinical Medicine, Vilnius, Lithuania; Dr M. Fleck, Department de Psychiatrie, Universite Legal, Porte Alegre, Brazil; K.-F. Leung, Hong Kong Hospital Authority, Kowloon, Hong Kong; Professor N. Von Steinbuchel, Centre of Neuro-gerontology psychology, Geneva University of Psychogeriatriecs, Geneva, Switzerland; Associate Professor E. Eiser, Celal Bayar University, Manisa, Turkey; Professor L. Schwartzmann, Department of
Medical Psychology, Uruguay; Dr R. Killian, Department of Psychiatry, University of Ulm, Germany; Dr S. Schmidt, University of Hamburg, Germany. Jiqian Fang and Mick Power have equally contributed to this work as the first authors.

References


