

Validation of the WHOQOL-BREF quality of life questionnaire for general use in New Zealand: confirmatory factor analysis and Rasch analysis

Christian U. Krägeloh · Paula Kersten ·
D. Rex Billington · Patricia Hsien-Chuan Hsu ·
Daniel Shepherd · Jason Landon · Xuan Joanna Feng

Accepted: 4 September 2012 / Published online: 15 September 2012
© Springer Science+Business Media B.V. 2012

Abstract

Purpose The present study validated the abbreviated version of the World Health Organization Quality of Life (WHOQOL-BREF) questionnaire for general use in New Zealand.

Methods A random postal sample from the national electoral roll was used, and 808 questionnaires were returned. Psychometric properties of the instrument were assessed, including tests of the four-domain factor structure using confirmatory factor analysis and Rasch analysis.

Results Goodness-of-fit from the confirmatory factor analysis were good, and the overall conclusion of the Rasch analysis supported the confirmatory factor analysis (CFA)

findings after dealing with problems of threshold ordering, local dependency, and differential item functioning (DIF). **Conclusions** The WHOQOL-BREF is valid for general use in New Zealand. In the future work, the WHOQOL-BREF domain scores should either be analyzed using non-parametric statistics or data should be fitted to the Rasch model to derive interval person estimates.

Keywords Quality of life · Questionnaire validation · WHOQOL-BREF · General population · New Zealand

Abbreviations

HRQOL	Health-related quality of life
WHOQOL	World Health Organization Quality of Life
CFA	Confirmatory factor analysis
DIF	Differential item functioning

C. U. Krägeloh (✉) · D. Rex Billington ·
P. H.-C. Hsu · D. Shepherd · J. Landon · X. J. Feng
Department of Psychology, Faculty of Health and Environmental
Sciences, Auckland University of Technology, North Shore
Campus, Private Bag 92006, Auckland 1142, New Zealand
e-mail: chris.krageloh@aut.ac.nz

D. Rex Billington
e-mail: rex.billington@aut.ac.nz

P. H.-C. Hsu
e-mail: gdpatricia@yahoo.com

D. Shepherd
e-mail: daniel.shepherd@aut.ac.nz

J. Landon
e-mail: jason.landon@aut.ac.nz

X. J. Feng
e-mail: joanna_xy@hotmail.com

P. Kersten
Person Centred Research Centre, School of Rehabilitation
and Occupation Studies, Faculty of Health and Environmental
Sciences, Auckland University of Technology, Private Bag
92006, Auckland 1142, New Zealand
e-mail: paula.kersten@aut.ac.nz

Introduction

Assessment of health-related quality of life (HRQOL) has become widespread recently, and among the most widely used questionnaires are the World Health Organization Quality of Life (WHOQOL) instruments [1]. Their appeal stems from the fact that they have been developed with international collaborations. Initially, 15 centers in 14 countries were involved in the development of the WHOQOL-100 [2, 3]. Thus, the WHOQOL can claim to possess broad cultural applicability [4, 5]. Its psychometric properties are acknowledged to be excellent [6].

As HRQOL measures are commonly taken in conjunction with other measures, participant burden is an important consideration. Therefore, the short version, the WHOQOL-BREF [5], has proven to be very popular [1]. Until now,

WHOQOL research in New Zealand [7–11] has used the Australian version of the WHOQOL-BREF [12], as no New Zealand version had been available. However, with the recent establishment of the New Zealand WHOQOL Group [1], more research will be conducted in New Zealand using the WHOQOL-BREF. It is timely that the suitability of this instrument is formally established for general use in New Zealand.

The purpose of the present study was to test the psychometric properties of the WHOQOL-BREF for use in New Zealand. Data were collected using a sample from the general population. Psychometric properties of the instrument were assessed, including tests of the four-domain factor structure using confirmatory factor analysis (CFA) and Rasch analysis.

Methods

Using the New Zealand national electoral roll, 3,000 questionnaires and self-addressed return envelopes were posted out randomly. A cover letter explained the purpose and procedure of the study. Four months after the mailout, 710 questionnaires had been returned (a response rate of 24 %). Since participants aged 18–30 were underrepresented (8.2 %) compared with national census figures (22.8 %) [13], additional data from this group were collected using purposive convenience sampling. Using community group networks, young adults were approached in Auckland (the largest city in New Zealand with a population of approximately 1,400,000) and Palmerston North (population of approximately 82,000). As a result, the total number of participants increased to 808, and the percentage of participants in the category 18–30 years increased to 18.4 %.

Participants completed the Australian version of the WHOQOL-BREF [12]. To control for order effects, half of the questionnaires presented items in a different order. The study had prior approval by the ethics committee of the authors' university. CFA was conducted with LISREL v. 8.80 [14] and Rasch analysis with RUMM2030 [15]. All remaining data analyses used SPSS v.18.0. The CFA used diagonally weighted least squares with polychoric correlations, as data were ordinal [16, 17]. The four-domain structure of the WHOQOL-BREF was tested, by allowing correlations between domains, but no correlations of error variance.

Each of the WHOQOL-BREF domains was tested against the polytomous partial credit Rasch model [18] to examine its internal construct validity [19]. Rasch analysis has been explained in detail elsewhere [20–23], so only the analytical concepts are briefly outlined in Table 1. Testlets were created if problems with response dependency or

differential item functioning (DIF) were found to test if bias is canceled out at the test level and also if they remove the dependency in the data [24]. Bonferroni corrections were used to allow for multiple testing.

Results

Of the 808 participants, 337 were male and 469 female (2 missing). The mean age was 49.69 years, with a standard deviation of 17.85. For their highest level of education, 50 % reported tertiary, 45 % secondary, and 3 % primary education. Fifty-eight percent were married, 18 % single, 8 % lived as married, 7 % divorced, 6 % widowed, and 2 % separated. To the question “Are you currently ill?” 140 participants answered “yes” and 656 answered “no.”

For the majority of WHOQOL-BREF items, kurtosis and skewness coefficients were within the acceptable range of -1.00 to 1.00 . Several items were slightly outside this range, but still within -1.10 to 1.10 . Item 23 (*condition of living place*) had a skewness and a kurtosis coefficient of -1.38 and 2.19 , respectively, and a mean of 4.30 ($SD = 0.84$). Item 25 (*transport*) had a mean of 4.06 ($SD = 0.93$) and a kurtosis coefficient of 1.18 .

Cronbach's alpha values were 0.91 for the overall scale, 0.80 for the physical domain (seven items), 0.82 for the psychological domain (six items), 0.71 for the social domain (three items), and 0.81 for the environment domain (eight items). All values were above 0.70 , and thus showed adequate internal consistency. Criterion-related validity was assessed by correlating item and domain scores with Items 1 (*global quality of life*) and 2 (*global health*). Items 1 and 2 were significantly correlated with all 24 remaining items ($P < 0.01$). All domain scores were significantly ($P < 0.01$) correlated with Item 1 (Pearson's r ranged from 0.45 to 0.60), as well as Item 2 (0.31 – 0.64). Domain mean scores and standard deviations are shown in Table 2.

The suitability of the factor structure of the WHOQOL-BREF was evaluated using three goodness-of-fit values [25]. Root mean square error of approximation was 0.072 , and thus above the <0.060 criterion for a good fit. The comparative fit index was 0.966 and thus above the >0.95 criterion for a good fit. A final goodness-of-fit index, the standardized root mean square residuals was 0.067 , thus indicating that the fit was good (criterion <0.080). Standardized factor loadings were all above 0.50 , except for Items 3 (*pain*) and 4 (*medication*), which had factor loadings of 0.48 and 0.36 , respectively. The correlations between the four factors ranged from 0.48 to 0.75 (Table 2).

A smaller sample was used to test the data against the Rasch model because large sample sizes can result in Type I errors, that is, falsely rejecting an item as not fitting the

Table 1 Fit to the Rasch model

Concept examined	Person fit	Item fit	Item threshold ordering ^a	Item differential item functioning (DIF) ^b	Response dependency	Overall fit to the Rasch model	Reliability index	Unidimensionality
Statistical test used	Mean fit residuals (SD); range	Mean fit residuals (SD); range; χ^2	Visual inspection	ANOVA	Residual correlations	Item-trait interaction χ^2	Person separation index (PSI)	Principal component analysis of the residuals
Acceptable values [19–22]	Mean close to zero and SD close to 1; range –2.5 to 2.5 χ^2 should be non-significant	Mean close to zero and SD close to 1; range –2.5 to 2.5 χ^2 should be non-significant	Must be ordered	Should be non-significant	<0.3 above the mean residual correlation [25]	Should be non-significant	≥ 0.70 allow for group comparisons; ≥ 0.85 for individual clinical use	The 95 % CI of the proportion of significant tests should include 5 % [23, 24]
Domains								
<i>Physical domain</i>								
Findings (only deviations reported)	Seventeen persons fit residual <–2.5	High item fit residual SD (1.947)	Response categories 3 and 4 of three items were combined (Q16, 17, 18)	Uniform DIF by the variables <i>age</i> (three items: Q4, 10, 16) and <i>unwell</i> (two items: Q16, 17)	Response dependency between three sets of items	✓	✓	✓
Solution: two testlets were created, combining two items that displayed DIF by <i>age</i> into one testlet and the items that displayed DIF by <i>unwell</i> in the other testlet	–0.350 (0.946); –3.97 to 2.08 (six persons fit residual <–2.5) ^c	–0.111 (1.504); –2.29 to 1.32; χ^2 non-significant	✓	✓	✓	$\chi^2 = 47.74$ (45); $P = 0.362$	0.79	4.3 % (0.2–6.4)
<i>Psychological domain</i>								
Findings (only deviations reported)	Twenty-four persons fit residual <–2.5	High item fit residual SD (2.333); one item high negative fit residual (–3.31)	✓	Uniform DIF by the variables <i>age</i> (two items: Q5, 26), <i>gender</i> (three items: Q6, 11, 19) and <i>education</i> (one item: Q7)	Response dependency between two items	Significant deviation ($P = 0.002$)	✓	✓
Solution: two testlets were created, combining two items that displayed DIF by <i>age</i> into one testlet and the items that displayed DIF by <i>gender</i> in the other testlet (with one final item remaining)	–0.455 (1.041); –4.54 to 1.55 (18 persons fit residual <–2.5) ^c	–0.355 (3.845); –3.86 to 3.76 ^d ; χ^2 non-significant	✓	✓	✓	$\chi^2 = 27.14$ (27); $P = 0.456$	0.79	2.4 % (0.3–4.5)
<i>Social domain</i>								

Table 1 continued

Concept examined	Person fit	Item fit	Item threshold ordering ^a	Item differential item functioning (DIF) ^b	Response dependency	Overall fit to the Rasch model	Reliability index	Unidimensionality
Findings (only deviations reported)	✓	One item significant $\chi^2 = 23.15$ (6), $P = 0.002$	Response categories 2 and 3 of two items were combined (Q21, 21)	Uniform DIF by the variables <i>age</i> (three items: Q20–Q22), <i>marital</i> (one item: Q20), <i>unwell</i> (one item: Q21)	Response dependency between two items	Significant deviation ($P = 0.004$)	Low PSI (0.54)	✓
Solution: One testlet was created, combining the DIF items and items with response dependency simultaneously	-0.472 (0.815); -4.53 to 1.23 (two persons fit residual <-2.5) ^c	-0.118 (3.360) ^d -2.49 to 2.26; χ^2 non- significant	✓	✓	✓	$\chi^2 = 14.41$ (13); $P = 0.346$	0.54	2.3 % (0.1–4.4)
<i>Environmental domain</i>								
Findings (only deviations reported)	Twenty-two persons fit residual <-2.5 ^c	✓	Response categories 2 and 3 of five items were combined (Q12, 13, 23, 24, 25)	Uniform DIF by the variable <i>version</i> (three items: (Q8, 24, 25)	Response dependency between two sets of items	✓	✓	✓
Solution: two testlets were created, combining two items that displayed DIF by <i>version</i> into the one testlet and items that displayed response dependency in the other testlet	-0.315 (1.036); -3.29 to 1.908 (11 persons fit residual <-2.5) ^c	0.346 (1.097);	✓	✓	✓	$\chi^2 = 50.61$ (45); $P = 0.262$	0.77	5.3 % (3.2–7.3)

^a Where disordered threshold and observed response categories were combined so that a more precise estimate of HRQOL could be obtained

^b DIF was examined with the following groups as the main factor: age, gender, educational attainment, marital status, whether someone was well or unwell, and WHOQOL version completed

^c A negative fit residual indicates that the person has a better than predicted fit to the Rasch model. As these persons do not “misfit” they were retained in the analysis

^d High fit residual and mean fit residual standard deviation here is due to the creation of a testlet as opposed to misfit

Table 2 Correlations between latent factors of the CFA

	Mean	Standard deviation	(1)	(2)	(3)	(4)
Physical (1)	27.39	4.83	–	0.74	0.48	0.66
Psychological (2)	22.46	3.88	0.74	–	0.74	0.75
Social (3)	11.56	2.42	0.48	0.74	–	0.67
Environmental (4)	31.48	4.78	0.66	0.75	0.67	–

All correlations were significant ($P < 0.01$)

Also shown are domain means and standard deviations

Rasch model [26]. Four hundred and twenty people were included in the analysis, a sample size large enough to have 99 % confidence that the estimated item difficulty is within $\pm 1/2$ logit of its stable value [27]. All participants who had self-identified as unwell ($n = 140$), as well as a random sample of 280 from the remaining participants were included. The findings are summarized in Table 1. For each of the domains, a small number of people fit the Rasch model better than expected, as indicated by the negative fit residuals smaller than -2.50 . However, as they did not misfit the model, they were retained in the analysis. Ten items had disordered thresholds and were successfully re-scored prior to further analysis. Each domain was unidimensional but included items that displayed uniform DIF by a range of variables. In addition, each domain had a few items that were dependent on one another. The creation of testlets for DIF or dependent items overcame these problems with subsequent item fit and overall fit to the Rasch model. The domains had acceptable reliability at the group level, except the social domain, which had a low Person Separation Index of 0.54.

Discussion

The present study evaluated the psychometric properties of the WHOQOL-BREF for use in the general population in New Zealand. Reliability and criterion-related validity were very good. A CFA suggested that the generally acknowledged four-domain structure provided a good to very good fit, with goodness-fit-values that were slightly better than those from similar analyses reported elsewhere [6, 28–30]. The overall conclusion of the Rasch analysis supported the CFA findings after dealing with problems of threshold ordering (in 10 items across three domains), local dependency (all domains), and DIF (all domains: 15/24 items).

Classical test theory assumes that thresholds are ordered, while in Rasch measurement theory this is specifically assessed [31]. Disordered thresholds are problematic and indicate that the response categories are not used

consistently along the latent trait. In our analysis, we were able to account for disordered thresholds by collapsing some item response categories, which is a common approach [32]. In one other study, disordered thresholds were found, though to a lesser extent [33]. By contrast, one study did not identify this issue at all [34], and others did not report on this [35, 36].

The assessment of local dependency has undergone changes recently. Current protocols are very strict, accepting a scale as locally independent if correlations between residuals are smaller than 0.30 [37]. Local dependency was only assessed in one other study [33] and was only an issue in the physical domain. The authors were able to resolve this by creating a testlet as was conducted here.

The present study found significant problems with DIF in 15 of 24 items. DIF was more prevalent than in other studies that formally assessed this [33, 34, 36]. Rather than deleting these items [36], we created testlets containing items that showed DIF in opposite directions (either statistically or visually observed) [24]. This led to satisfactory fits to the Rasch model.

Our analysis differed from that by Wang et al. [36] who used a multi-dimensional Rasch model. In the multi-dimensional model, domains are simultaneously calibrated and the correlations between traits are taken into account [38]. The disadvantage of this approach is that the raw scores are no longer sufficient statistics for the Rasch derived person estimates since their estimates on one domain are dependent upon the other three domains as well. Our findings also differ from a Danish study, which established that the WHOQOL-BREF did not fit the Rasch model but that it did fit a two-parameter model [35]: This is an item response theory model that allows different discrimination of items in a scale. Problems with this approach have been widely documented and essentially concern the model's inability to provide a separate estimation of item and person parameters, which highlights fundamental measurement problems [39].

The following limitations need to be acknowledged. Firstly, the New Zealand electoral roll was used to collect a sample of the general population, thus equating general population with the population of people enrolled to vote. Typically, more than 90 % of the eligible voting population (which includes people with New Zealand citizenship or permanent residency) is enrolled to vote, although this proportion is lower for people aged 18–24. Secondly, the extent of self-selection bias is likely to have been strong. The low response rate of younger participants prompted collection of further data using purposive sampling, but a gender bias still remained, with 58 % of respondents being female. No data on ethnicity were collected, and it therefore cannot be determined to what extent the sample

represents the ethnic mix of the New Zealand population. The present study was conducted 1 year after the 2008 national elections, and people who had moved house since the elections would have received the questionnaire at their old address. This may explain why the response rate was particularly low for young people, who may be prone to moving house more frequently than others. Due to budget constraints and confidentiality requirements, no incentives, sweepstakes, reminder letters, etc. could be offered, which would have increased the response rate [40].

Unlike the original validation studies of the WHOQOL-BREF [5], the present study did not purposively collect data from individuals with identified health issues as it was intended to validate the WHOQOL-BREF for use in the general population. This intention was a considering factor during the interpretation of the factor structure of the instrument. A frequently used cutoff criterion for factor analysis is that factor loadings need to be above 0.30, although 0.40 appears to be more frequently used. In our case, the factor loading of 0.36 for Item 4 (*medication*) was therefore marginal. However, since the WHOQOL-BREF is a well-established measure that has been used worldwide, we believed a laxer criterion to be justifiable. Additionally, the item fit the Rasch model and was thus retained in the analysis. Our own previous work [41] and that of others [42] found that this item, as well as Item 3 (*pain*), did not perform well in samples of young university students with a low proportion of self-reported health problems. The fact that the present study finds similar (albeit less pronounced) problems with these items could reflect the fact that a large proportion of respondents in our sample also did not have health issues.

To summarize, the WHOQOL-BREF is suitable for use in New Zealand with samples from the general population. As a result of the collapsing of some of the response categories in our analysis, the total achievable domain scores have changed. Consequently, it is not possible to provide a straightforward ordinal-to-interval scores conversion table. Future work should either analyze WHOQOL-BREF domain scores using non-parametric statistics or data should be fitted to the Rasch model to derive interval person estimates.

Acknowledgments This research was funded by the Faculty of Health and Environmental Sciences of AUT University.

References

1. Billington, R., Landon, J., Krägeloh, C. U., & Shepherd, D. (2010). The New Zealand WHOQOL Group. *New Zealand Medical Journal*, 123(1315), 65–70.
2. Power, M., Bullinger, M., Harper, A., & The WHOQOL Group. (1999). The World Health Organization WHOQOL-100: Tests of the universality of quality of life in 15 different cultural groups worldwide. *Health Psychology*, 18(5), 495–505. doi:10.1037/0278-6133.18.5.495.
3. WHOQOL Group. (1995). The World Health Organization quality of life assessment (WHOQOL): Position paper from the World Health Organization. *Social Science and Medicine*, 41(10), 1403–1409.
4. Bowden, A., & Fox-Rushby, J. A. (2003). A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe, the Middle East, South America. *Social Science and Medicine*, 57(7), 1289–1306. doi:10.1016/S0277-9536(02)00503-8.
5. WHOQOL Group. (1998). Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychological Medicine*, 28, 551–558.
6. Skevington, S. M., Lotfy, M., & O'Connell, K. A. (2004). The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. *Quality of Life Research*, 13(2), 299–310. doi:10.1023/B:QURE.0000018486.91360.00.
7. Chai, P. P. M., Krägeloh, C. U., Shepherd, D., & Billington, R. (2012). Stress and quality of life in international and domestic university students: Cultural differences in the use of religious coping. *Mental Health, Religion and Culture*, 15(3), 265–277. doi:10.1080/13674676.2011.571665.
8. Henning, M. A., Krägeloh, C. U., Sameshima, S., Shepherd, D., Shepherd, G., & Billington, R. (2011). Access to New Zealand sign language interpreters and quality of life of the deaf: A pilot study. *Disability and Rehabilitation*, 33(25–26), 2559–2566. doi:10.3109/09638288.2011.579225.
9. Hsu, P. H.-C., Krägeloh, C. U., Shepherd, D., & Billington, R. (2009). Religion/spirituality and quality of life of international tertiary students in New Zealand: An exploratory study. *Mental Health, Religion, and Culture*, 12(4), 385–399. doi:10.1080/13674670902752920.
10. Keogh, J. W. L., Shepherd, D., Krägeloh, C. U., Ryan, C., Masters, J., Shepherd, G., et al. (2010). Predictors of physical activity and quality of life in New Zealand prostate cancer survivors undergoing androgen-deprivation therapy. *New Zealand Medical Journal*, 123(1325), 20–29.
11. Taylor, W. J., Myers, J., Simpson, R. T., McPherson, K. M., & Weatherall, M. (2004). Quality of life of people with rheumatoid arthritis as measured by the World Health Organization Quality of Life instrument, short form (WHOQOL-BREF): Score distributions and psychometric properties. *Arthritis and Rheumatism*, 51(3), 350–357. doi:10.1002/art.20398.
12. Murphy, B., Herrman, H., Hawthorne, G., Pinzone, T., & Evert, H. (2000). *Australian WHOQoL instruments: User's manual and interpretation guide*. Melbourne: Australian WHOQoL Field Study Centre.
13. Statistics New Zealand (2006). <http://www.stats.govt.nz/Census/2006CensusHomePage/classification-counts-tables/about-people/age.aspx>.
14. Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modelling with the SIMPLIS command language*. Hillsdale, NJ: Erlbaum Associates.
15. Andrich, D., Sherridan, B., & Luo, G. (2009). *RUMM2030*. Perth: RUMM Laboratory Pty Ltd.
16. Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466–491. doi:10.1037/1082-989X.9.4.466.
17. Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, 24(4), 387–404. doi:10.1007/BF00152012.

18. Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests (revised and expanded ed.)*. Chicago: The University of Chicago Press.
19. Goodwin, L. D., & Goodwin, W. L. (1991). Focus on psychometrics. Estimating construct validity. *Research in Nursing and Health, 14*(3), 235–243.
20. Andrich, D. (1988). *Rasch models for measurement series: Quantitative applications in the social sciences no. 68*. London: Sage Publications.
21. Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model. Fundamental measurement in the human sciences*. London: Lawrence Erlbaum Associates.
22. Kersten, P., & Kayes, N. M. (2011). Outcome measurement and the use of Rasch analysis, a statistics-free introduction. *New Zealand Journal of Physiotherapy, 39*(2), 92–99.
23. Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care and Research, 57*(8), 1358–1362. doi: [10.1002/art.23108](https://doi.org/10.1002/art.23108).
24. Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement, 24*(3), 185–202. doi: [10.1111/j.1745-3984.1987.tb00274.x](https://doi.org/10.1111/j.1745-3984.1987.tb00274.x).
25. Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453. doi: [10.1037/1082-989X.3.4.424](https://doi.org/10.1037/1082-989X.3.4.424).
26. Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology, 8*, 33.
27. Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7*(4), 328.
28. Jaracz, K., Kalfoss, M., Górna, K., & Bączyk, G. (2006). Quality of life in Polish respondents: Psychometric properties of the Polish WHOQOL-BREF. *Scandinavian Journal of Caring Sciences, 20*(3), 251–260. doi: [10.1111/j.1471-6712.2006.00401.x](https://doi.org/10.1111/j.1471-6712.2006.00401.x).
29. Yao, G., Wang, J.-D., & Chung, C.-W. (2007). Cultural adaptation of the WHOQOL questionnaire for Taiwan. *Journal of the Formosan Medical Association, 106*(7), 592–597. doi: [10.1016/S0929-6646\(07\)60012-6](https://doi.org/10.1016/S0929-6646(07)60012-6).
30. Xia, P., Li, N., Hau, K.-T., Liu, C., & Lu, Y. (2012). Quality of life of Chinese urban community residents: A psychometric study of the mainland Chinese version of the WHOQOL-BREF. *BMC Medical Research Methodology, 12*, 37. doi: [10.1186/1471-2288-12-37](https://doi.org/10.1186/1471-2288-12-37).
31. Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573. doi: [10.1007/BF02293814](https://doi.org/10.1007/BF02293814).
32. Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology, 46*(1), 1–18. doi: [10.1348/014466506X96931](https://doi.org/10.1348/014466506X96931).
33. da Rocha, N. S., & Fleck, M. P. d. A. (2009). Validity of the Brazilian version of WHOQOL-BREF in depressed patients using Rasch modelling. *Revista De Saúde Pública, 43*(1), 147–153. doi: [10.1590/S0034-89102009000100019](https://doi.org/10.1590/S0034-89102009000100019).
34. Liang, W.-M., Chang, C.-H., Yeh, Y.-C., Shy, H.-Y., Chen, H.-W., & Lin, M.-R. (2009). Psychometric evaluation of the WHOQOL-BREF in community-dwelling older people in Taiwan using Rasch analysis. *Quality of Life Research, 18*, 605–618. doi: [10.1007/s11136-009-9471-5](https://doi.org/10.1007/s11136-009-9471-5).
35. Noerholm, V., Groenvold, M., Watt, T., Bjorner, J. B., Rasmussen, N.-A., & Bech, P. (2004). Quality of life in the Danish general population—normative data and validity of WHOQOL-BREF using Rasch and item response theory models. *Quality of Life Research, 13*(2), 531–540. doi: [10.1023/B:QURE.0000018485.05372.d6](https://doi.org/10.1023/B:QURE.0000018485.05372.d6).
36. 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*(4), 607–620. doi: [10.1007/s11136-005-4365-7](https://doi.org/10.1007/s11136-005-4365-7).
37. Marais, I., & Andrich, D. (2008). Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model. *Journal of Applied Measurement, 9*(2), 105–124.
38. Wang, W.-C., Chen, P.-H., & Cheng, Y.-Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods, 9*(1), 116–136. doi: [10.1037/1082-989X.9.1.116](https://doi.org/10.1037/1082-989X.9.1.116).
39. Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care, 42*(1), I7–I16. doi: [10.1097/01.mlr.0000103528.48582.7c](https://doi.org/10.1097/01.mlr.0000103528.48582.7c).
40. Edwards, P., Roberts, I., Clarke, M., DiGiuseppe, C., Pratap, S., Wentz, R., et al. (2002). Increasing response rate to postal questionnaires: Systematic review. *BMJ, 324*, 1–9. doi: [10.1136/bmj.324.7347.1183](https://doi.org/10.1136/bmj.324.7347.1183).
41. Krägeloh, C. U., Henning, M. A., Hawken, S. J., Zhao, Y., Shepherd, D., & Billington, R. (2011). Validation of the WHOQOL-BREF quality of life questionnaire for use with medical students. *Education for Health, 24*(2), 1–5.
42. Li, K., Kay, N. S., & Nokkaew, N. (2009). The performance of the World Health Organizations's WHOQOL-BREF in assessing the quality of life of Thai college students. *Social Indicators Research, 90*(3), 489–501. doi: [10.1007/s11205-008-9272-1](https://doi.org/10.1007/s11205-008-9272-1).