RESEARCH

Short-form development of the specific module of the QLICD-CRF(V2.0) for assessing the quality of life of patients with chronic renal failure

Zhengqin Xiao[1†], Yuxi Liu[1†], Daniel Yee-Tak Fong[2], Xinping Huang[3], Min Weng[4] and Chonghua Wan[1*]

Abstract

Background: A short instrument would enhance the viability of a study. Therefore, we aimed to shorten the specific module (SPD-10) of the Quality of Life Instrument for Chronic Diseases - Chronic Renal Failure (QLICD-CRF) for assessing the quality of life of patients with chronic renal failure.

Methods: The 10-item SPD-10 was self-administered to 164 patients with chronic renal failure. A shortened form was first obtained by a tandem use of the classical test theory (CTT), the generalizability theory (GT), and the item response theory (IRT). In addition, we also shortened the SPD-10 by the Optimal Test Assembly (OTA).

Results: Both the tandem use of GT, CTT and IRT, and the OTA derived the same 7-item shortened version (SPD-7). It included items CRF1, CRF2, CRF3, CRF4, CRF6, CRF8, and CRF9 of the SPD-10. The SPD-7 had a Cronbach alpha of 0.78. The correlation coefficients of its total and factor scores with those of the SPD-10 were 0.96 and 0.98, respectively. Confirmatory factor analysis confirmed the unidimensional structure of the SPD-7, with the comparative fit index=0.96, the Tucker-Lewis index=0.94, and the root mean square error of approximation=0.09.

Conclusion: The short-form SPD-7 is reliable and valid for assessing the impact of clinical symptoms and side effects on the quality of life of patients with chronic renal failure. It is an efficient option without compromising the measurement performance of the SPD-10.

Keywords: Scale reduction, Generalizability theory, Classical test theory, Item response theory, Optimal test assembly

Background

Chronic renal failure (CRF) is an irreversible deterioration of renal function, which has become a major public health problem worldwide [1]. Its clinical manifestations,

†Zhengqin Xiao and Yuxi Liu are the first co-author with the same contributions.

*Correspondence: wanchh@hotmail.com

[1] School of Humanities and Management, Research Center for Quality of Life and Applied Psychology, Key Laboratory for Quality of Life and Psychological Assessment and Intervention, Guangdong Medical University, Dongguan 523808, China
Full list of author information is available at the end of the article

such as anemia, uremia, hypertension, anorexia, and pruritus, can cause severe physical discomfort to the patients [2]. Moreover, the long-term treatment of CRF renders patients prone to the development of depression, anxiety, and other psychiatric problems [2, 3]. Consequently, CRF patients may have deteriorated quality of life (QOL).

To meet the urgent need for proper evaluation of the QOL of CRF patients in China, the Quality of Life Instrument for Chronic Diseases-Chronic Renal Failure (QLICD-CRF) was developed [4, 5]. The latest version, the 38-item QLICD-CRF V2.0, included a 28-item general module (QLICD-GM V2.0) that applies to patients

Xiao *et al. BMC Medical Research Methodology*   (2022) 22:289

Page 2 of 8

with any chronic diseases, and a 10-item module specific for CRF patients (SPD-10). The instrument has demonstrated good reliability and validity for measuring the QOL of Chinese CRF patients [4, 5]. However, the instrument is lengthy, which adds a substantial administrative burden and reduces data quality, especially when it is applied in outpatient or community settings. Therefore, it is important to obtain a short-form without compromising its reliability and validity.

There have been several common instrument shortening approaches, including the Generalizability Theory (GT), the Classical Test Theory (CTT) and the Item Response Theory (IRT). The GT determines an appropriate number of items of an instrument by estimating the dependability of scores with the number of items in different forms [6–8]. However, GT does not select items in a shortened form. Thus, the GT method is often used with an item selection strategy. The CTT and IRT methods have been the most popular item selection methods. A CTT model assumes a subject's observed score is a composite of a true score and measurement error [9–11]. In contrast, IRT directly models an item's response on its corresponding latent trait and derives measurement properties that are sample independent [12, 13]. To take advantage of the strengths of the three methods, a tandem approach can be adopted [14, 15].

More recently, an alternative scale reduction approach, called the Optimal Test Assembly (OTA), was proposed [16, 17]. The OTA is a mixed-integer programming procedure that uses an estimated IRT model to select a subset of items that best satisfies a pre-specified set of constraints. It can be described as a two-stage process. Stage one searches for the set of items that maximizes the test information for each fixed number of items. In stage two, the minimal set of items that meets the pre-specified criteria would be taken as the optimal short-form [18]. Examples of the criteria are having a Cronbach alpha of at least 95% of that of the full-form and a correlation of at least 0.95 with the full-form. Thus, given the same set of criteria, the short-form obtained by OTA is highly reproducible [16, 17, 19].

Consequently, we attempted to develop a short-form of the SPD-10 domain in the QLICD-CRF to facilitate effective assessment of the QOL in CRF patients, by a tandem use of the three test theories, and the OTA.

Methods
Sample
We recruited subjects aged 14 years or older who were (1) diagnosed with CRF; and (2) able to read Simplified Chinese. Subjects who had disturbance of consciousness, cognitive impairment, or other mental illness were excluded. Patients were recruited at the Department of Nephrology of the First Affiliated Hospital of the Kunming Medical University in China.

Eligible patients were explained the study procedures before they were invited to sign a consent form. A total of 164 patients with CRF consented to their study participation. The study protocol and consent form were approved by the Institutional Review Board of the investigators' affiliated institution.

Measures
The SPD-10 included 10 items on the frequency of symptoms or severity of side effects in the past week, with each item responding from 1 ("extraordinary") to 5 ("none at all"). Table 1 shows the 10 items of the SPD-10. The total score was standardized in the range from 0 to 100. A higher score indicated milder clinical symptoms and better quality of life [4, 5, 20].

Statistical analysis
Unidimensionality of the SPD-10
The unidimensionality of the SPD-10 was assessed by confirmatory factor analysis (CFA).

Specifically, we fitted a one-factor model and obtained the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). If the CFI and TLI $\geq$ 0.90 and the RMSEA $\leq$ 0.01, we concluded the one-factor model of the SPD-10 was acceptably and thus unidimensional [21].

Scale reduction by tandem use of GT, CTT and IRT
The GT encompassed the Generalizability Study (G-study) and the Decision Study (D-study). The G-study estimated the variance components of the measurement error of the test score due to various sources by using the analysis of variance (ANOVA) [11, 22]. The D-study accounted all sources of error when assessing

Table 1 Item information of the SPD-10

Item	Content description
CRF1	Have you had chest tightness, shortness of breath, or cough with pink foam sputum
CRF2	Have you had feel dizzy or pale
CRF3	Have you had edema of eyelids or face or lower limbs
CRF4	Have you had cramps in your limbs
CRF5	Have you had skin itchiness
CRF6	Have you had sleep wrong (do not sleep at night, want to sleep during the day)
CRF7	Have you urinated a lot at night
CRF8	Have you had muscle or joint pain
CRF9	Have you had decreased urine output
CRF10	Have you had constipation

Xiao *et al. BMC Medical Research Methodology*   (2022) 22:289

Page 3 of 8

the generalizability of the test score for a fixed number of items [10, 23].

Specifically, in the G-study, we considered the SPD-10 test score as a fixed facet and the 10 items as random by assuming they were randomly selected indicators from an underlying pool. Using the ANOVA, the variance components due to persons, items, and person by item interaction were estimated. In the D-study, we obtained the generalizability coefficient (G), the reliability index (ϕ), the relative error variance (σ_δ^2) and the absolute error variance (σ_Δ^2) for a fixed number of items from 3 to 10. The smallest number of items that had G and ϕ > 0.70, and σ_δ^2 and σ_Δ^2 < 0.20 was considered the required number of items in the short-form [10].

Then, we selected the items that satisfied the highest number of criteria based on estimates obtained from the CTT and IRT. The criteria were (1) an item-total correlation coefficient $\geq$ 0.4; (2) a reduced Cronbach alpha coefficient after deleting the item; (3) a corrected item-total correlation (CITC) $\geq$ 0.4 [24, 25]; (4) a factor loading $\geq$ 0.4 in the one-factor CFA [22, 26]; (5) an item discrimination parameter (α) estimate > 0.5; (6) estimates of all the threshold parameters (β) fell between -4 and 4 [27]; and (7) an average item information $\geq$ 0.5, which was roughly equivalent to the classical reliability estimate of 0.8.

When multiple sets of items satisfied the aforementioned criteria, the one with the highest average test information was selected [16].

Scale reduction by OTA

We have also used OTA to search for an optimal reduced set of items without compromising the precision, reliability, and concurrent validity. First, for a fixed number of items from 3 to 10, we used a branch and bound algorithm to search for the set of items that maximized the test information function (TIF) anchored at five points (-3, -1, 0, 1, 3) [18, 19]. The TIF was obtained based on a GPCM. We considered a minimum of three items to ensure a single factor model was identifiable [28]. Thus, eight candidate short-forms were obtained. Second, the final short-form was identified based on the following criteria: (1) a Cronbach alpha at least 95% of that of the SPD-10; and (2) a correlation with both the total and factor scores of the SPD-10 of at least 0.95. These criteria were set to ensure that the short-form would closely resemble the test precision, reliability, and validity of the full-form. The shortest form that fulfilled the criteria was chosen as the final shortened version.

Statistical software

The CFA was conducted in Mplus7 [29]. The GT analysis was conducted using mGENOVA [30, 31]. The CTT analysis was conducted in SPSS 24.0. The GPCM was

fitted using the ltm package in RStudio4.0.5 [32]. The OTA analysis was conducted using the lpSolveAPI package in RStudio4.0.5 [33]. All other data analyses were analyzed using RStudio4.0.5 [34].

Results

Characteristics of the sample and the SPD-10

A total of 164 hospitalized patients with CRF consented to participate in our study. Their mean age was 45.6 (standard deviation [SD] = 14.9, range = 14 to 81 years). Of which, 97 (59.1%) were male, 128 (78.0%) were ethnic Han, and 138 (84.1%) were married. Eighty-four (51.2%) patients had poor economic income, 62 (37.8%) were farmers, and 48 (29.3%) attended primary school only.

Unidimensionality of the SPD-10

The one-factor model of the SPD-10 was adequate in the CFA, with the CFI = 0.95, TLI = 0.94, and RMSEA = 0.07. This indicated the SPD-10 was unidimensional.

Scale reduction by a tandem use of GT, CTT and IRT

Figure 1 depicts the indices of the D-study in the GT analysis by each number of items of the SPD-10. A minimum of seven items were required to have the G and ϕ above 0.70, and the σ_δ^2 and σ_Δ^2 lower than 0.20. Table 2 shows the scaling properties of the 10 items of the SPD-10 by the CTT and IRT analyses. Except for item CRF7, all items had the corrected item-total correlation, Cronbach alpha, item-total correlation, and factor loadings met the predetermined standards. Most items had an average item information lower than 0.50. Overall, the seven items that satisfied the highest number of criteria were CRF1, CRF2, CRF3, CRF4, CRF6, CRF8, and CRF9.

Scale reduction by OTA

Table 3 shows the selected items for each of the eight candidate short-forms, and the corresponding Cronbach alphas and concurrent validity with the SPD-10. The shortened versions with at least seven items satisfied our pre-specified criteria on reliability and concurrent validity. Figure 2 shows the individual item information curves for the SPD-10. Those for the three removed items were flat. The 7-item version, SPD-7, which included CRF1, CRF2, CRF3, CRF4, CRF6, CRF8, and CRF9, was identical to the one derived by the tandem use of GT, CTT and IRT.

Reliability of the SPD-7

Table 4 reports the reliability of the SPD-7. The Cronbach alpha of the SPD-7 was 0.78, and the alpha after deleting one item ranged from 0.74 to 0.77. The split-half reliability was 0.75. Compared with the SPD-10, the number of items in the short-form decreased by 30%, with only a
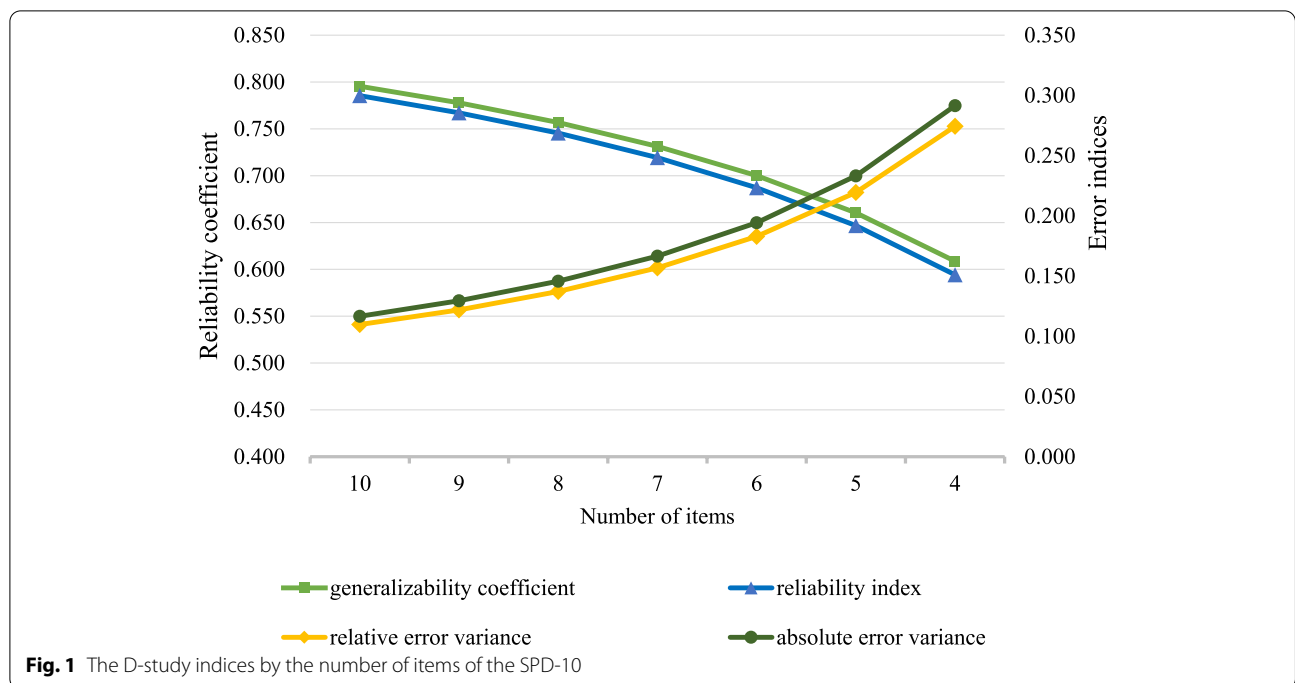
Xiao *et al. BMC Medical Research Methodology*   (2022) 22:289

Page 4 of 8

Fig. 1 The D-study indices by the number of items of the SPD-10

Table 2 Scaling information of the SPD-10 items based on the classical test theory (CTT) and item response theory analyses

Item	CTT				Generalized Partial Credit Model						Number of criteria satisfied
	①	②	③	④	①	②	③	④	⑤	⑥	
CRF1	0.563**	0.781	0.435	0.566	0.704	-2.061	-0.731	-2.360	-0.049	0.425	6
CRF2	0.639**	0.773	0.511	0.639	0.789	-1.743	-0.280	-1.199	1.858	0.489	6
CRF3	0.674**	0.770	0.531	0.644	0.707	-0.921	-0.177	-0.862	0.615	0.480	6
CRF4	0.571**	0.777	0.476	0.583	0.661	-1.055	-1.790	-1.351	0.850	0.399	6
CRF5	0.582**	0.777	0.477	0.569	0.572	-0.488	-1.669	-1.461	0.168	0.353	6
CRF6	0.620**	0.768	0.558	0.645	0.987	-1.681	-0.915	-1.594	0.190	0.682	7
CRF7	0.338**	0.808	0.192	0.238	0.194	-2.969	-2.959	-0.955	3.232	0.050	1
CRF8	0.609**	0.774	0.500	0.579	0.686	-1.755	-1.207	-0.831	1.453	0.402	6
CRF9	0.675**	0.767	0.549	0.661	0.849	-0.831	-0.921	-0.414	1.227	0.581	7
CRF10	0.585**	0.779	0.456	0.523	0.522	-1.094	-1.381	-0.912	0.268	0.304	6

CTT: ① Item-total correlation, ② Cronbach alpha, ③ Corrected item-total correlation, ④ Factor loading; GPCM: ① a, ②~⑤ β_1~β_4, ⑥ Average information. **$P<0.01$

2.1% reduction in alpha and 1.4% reduction in the split-half reliability.

Construct validity and Concurrent validity of the SPD-7

A one-factor model was fitted to the SPD-7 in a CFA. The model fit indices were TLI = 0.94 (> 0.90), CFI = 0.96 (> 0.90), and RMSEA = 0.09 (< 0.1), indicating acceptable model fit.

Moreover, the correlation of the total score of the SPD-7 with that of the full-form was $r=0.96$. The correlation of the factor scores between the full and the shortened versions was $r=0.98$. The two correlation coefficients were high [35], indicating that the SPD-7 well resembled the full-form.

Test Information of the SPD-7

Figure 3 shows the test information over (-3, 3) of the SPD-10, and that of the SPD-7. The test information of the SPD-7 closely resembled that of the SPD-10.

Discussions

We made the first attempt to develop a short-form of the SPD-10 for CRF patients by two approaches. The first approach utilized the GT, CTT, and IRT

Xiao *et al. BMC Medical Research Methodology*   (2022) 22:289

Page 5 of 8

Table 3 The selected items for short form based on OTA and their psychometric properties

Number of items	Items selected										Reliability	Concurrent validity with the SPD-10	
	CRF1	CRF2	CRF3	CRF4	CRF5	CRF6	CRF7	CRF8	CRF9	CRF10	Cronbach alpha	Correlation of the factor scores	Correlation of the total scores
3		√				√			√		0.648	0.892**	0.851**
4		√	√			√			√		0.721	0.923**	0.886**
5		√	√			√		√	√		0.733	0.952**	0.929**
6	√	√	√			√		√	√		0.758	0.962**	0.938**
7	√	√	√	√		√		√	√		0.778	0.980**	0.957**
8	√	√	√	√	√	√		√	√		0.797	0.990**	0.971**
9	√	√	√	√	√	√		√	√	√	0.808	0.999**	0.987**
10	√	√	√	√	√	√	√	√	√	√	0.795	1.000**	1.000**

**P<0.01

Fig. 2 Item Information Curves for the SPD-10. Each number on the curves corresponds to the item number
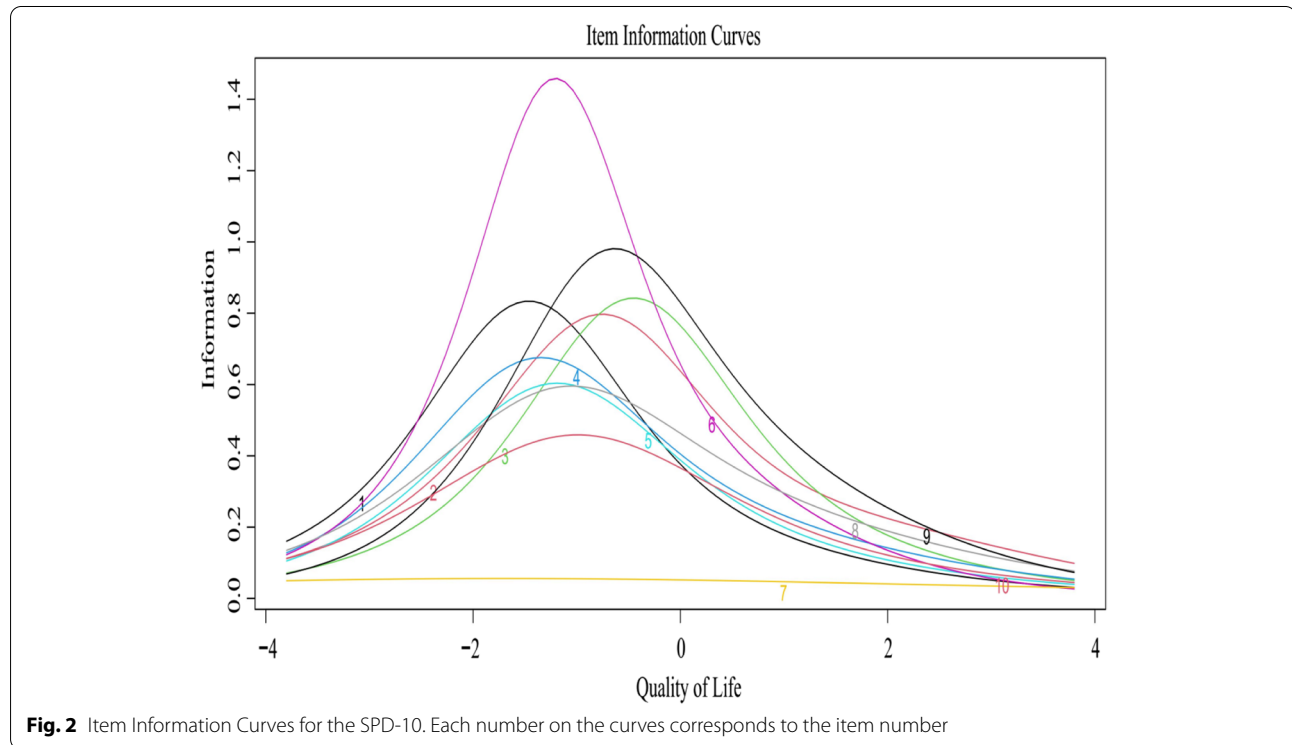
Table 4 Comparison of reliability indexes between SPD-7 and SPD-10

Scale	Number of items		Cronbach alpha		Split half reliability		Score distribution		
	Value	Variance (%)	Value	Variance (%)	Value	Variance (%)	Score	Median (*Q1, Q3*)	Mean
SPD-10	10	—	0.80	—	0.76	—	17.50 ~ 95.00	65.00 (50.00 ~80.00)	64.51
SPD-7	7	30.00	0.78	2.14	0.75	1.44	10.71 ~ 100.00	65.00 (50.89~78.57)	64.77

SPD-10:a chronic renal failure-specific module; SPD-7: a 7-item shortened form obtained. *Q1:* the item score is the 25% from small to large. *Q3:* the item score is the 75% from small to large

Fig. 3 Test Information Curves for the SPD-10 (red line) and the SPD-7 (green line)
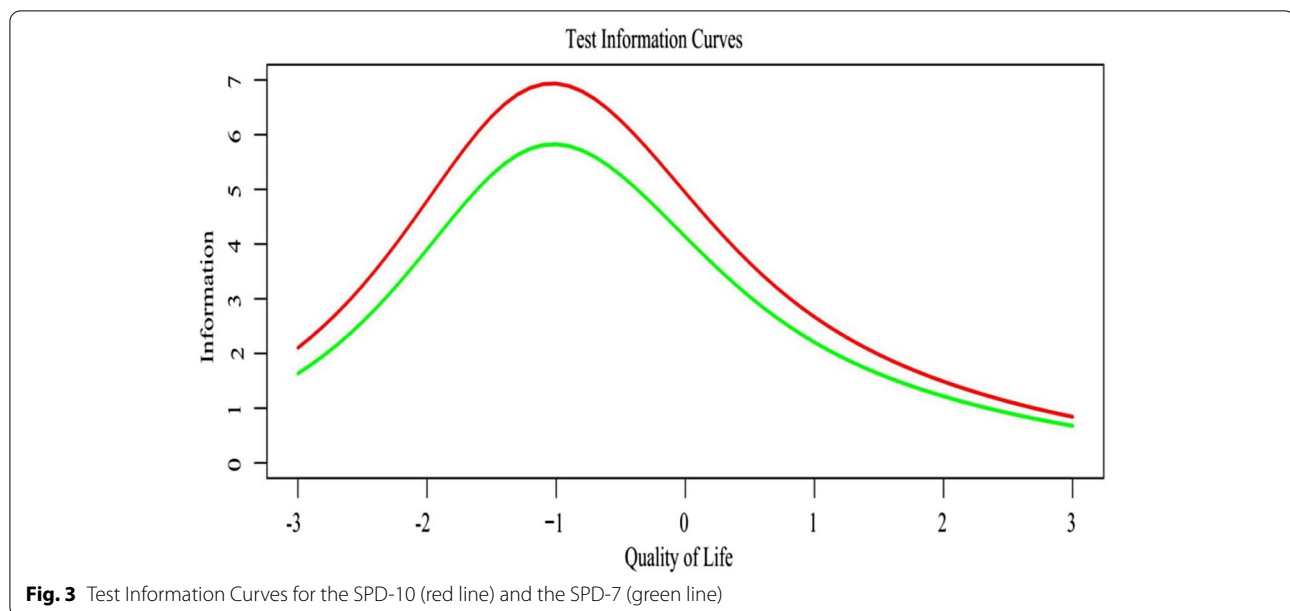
simultaneously. The GT analysis (D-study) showed seven items were adequate, with G and ϕ >0.7, and σ_δ^2 and σ_Δ^2 <0.2 [10]. The availability of such thresholds enhanced the objectivity in determining the required number of items. The CTT and IRT analysis selected seven items based on those with the higher average item information, and thus contributed to a more comprehensive perspective in the scale reduction process. In the second approach, we showed how the OTA could be used to develop a short-form of the SPD-10 based on pre-specified objective criteria. Coincidentally, both the OTA and the tandem use of GT, CTT and IRT derived the same short-form, the SPD-7.

Both approaches removed items CRF7, CRF5 and CRF10. In particular, item CRF7 "Have you urinated a lot at night?" performed poorly in discrimination, validity and average information. This can be due to the fact that nocturia is one of the earliest manifestations of CRF, which most patients would ignore [36]. The other two items: CRF5 "Have you had skin itchiness?" and CRF10 "Have you had constipation?" had the lowest IIF curves, indicating the items did not add to the precision for measuring QOL [37]. Indeed, a previous study showed that most patients with CRF had pruritus [38], which is also observed in our survey. When most patients have frequent pruritus, it may not add information to distinguish CRF patients with different QOL due to pruritus. Similarly, item CRF10 on constipation also had very low information; thus, it was also not included. Thus, although the three items cover symptoms relevant to CRF patients and their removal may be seen as reducing the content validity, their high

prevalence did not warrant their inclusion for discriminating the QOL of CRF patients.

Reliability analysis showed that the SPD-7's Cronbach alpha and split-half reliability were 0.78 and 0.75, respectively, corresponding to only 2.14% and 1.44% reduction of those of the full-form. Since the commonly accepted minimum level of reliability is 0.70 [39], the SPD-7 has acceptable reliability. Moreover, the CFA confirmed the validity of the one-factor model of the SPD-7, assuring its construct validity [21]. In addition, the total and factor scores of the SPD-7 were highly associated with those of the SPD-10, implying its good concurrent validity. Furthermore, the test information curves of the SPD-7 and the SPD-10 closely resembled each other, indicating that the short-form holds similar precision to measure QOL of CRF patients at the same latent trait level [40].

Overall, when considering internal consistency, correlation of the total and factor scores, construct validity and test information, the short-form SPD-7 showed good performance without compromising the performance of the original form. Therefore, SPD-7 is a viable alternative that can improve the feasibility of assessing QOL in CRF patients.

Generally speaking, the tandem approach of using the GT, CTT and IRT selected items with better quality from different aspects, and the OTA approach selected the short-form that best resembles the full-form in selected measurement properties. Specifically, the tandem use of the GT, CTT, and IRT involved a subjective selection of items based on their IIF despite objectively determining the number of items. Thus, the approach may result in different shortened versions, when used

Xiao *et al. BMC Medical Research Methodology*   (2022) 22:289

Page 7 of 8

by different researchers. On the other hand, the OTA allows objective determination of items based on pre-specified criteria. Nevertheless, both approaches arrived at the same short-form in our application and thus, the SPD-7 should be the most robustly derived.

There were several limitations worth noting. First, the sample size of this study was not very large, which may affect some results, e.g., the factor analysis and the IRT. Indeed, the IRT generally requires a large sample size for adequate stability. However, a study that investigates the effects of test length and sample size on IRT parameters found that a sample size as small as 150 could still be used with tests of 10, 20, or 30 items with enough accuracy [41]. Thus, with only 10 items in SPD-10, our sample of 164 would remain reasonable for the IRT analysis. Second, we included inpatients only, which may affect the generalizability of the results. Third, we had not evaluated the short-form in an independent sample. Future studies evaluating the SPD-7 in outpatients and inpatients through external data would be desirable.

Conclusions

The SPD-7 can be used for assessing the impacts of clinical symptoms and side effects on QOL in CRF patients, with reliability and validity as good as its full-form. It is an efficient option, especially when used with the general module QLICD-GM for a more comprehensive QOL assessment.

Abbreviations

GT: Generalizability theory; CTT: Classical test theory; IRT: Item response theory; OTA: Optimal test assembly; CRF: Chronic renal failure; QOL: Quality of life; QLICD-CRF: Quality of Life Instruments for Chronic Disease-chronic renal failure; SPD-10: a 10-item specific module for chronic renal failure; SPD-7: a 7-item short form of specific module for chronic renal failure; GPCM: Generalized partial credit model.

Acknowledgements

When carrying out this research project, we have received substantial assistance from many staff members of the First Affiliated Hospital of Kunming Medical University, we sincerely acknowledge all the support.

Authors' contributions

CHW designed the study. XPH, MW, performed the data collection. ZQX, YXL, DF performed data analyses and drafted the manuscript. CHW revised the manuscript deeply. All authors contributed to interpreting the data, and have read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (71373058, 81460519), Science and Technological Planning Program of Guangdong Province (2013B021800074).

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study protocol and the informed consent form were approved by the IRB (institutional review board) of Kunming Medical University (30860248) and Guangdong Medical University (PJ2013037). The study participation was voluntary and the respondents provided written informed consent before their participation.

All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Humanities and Management, Research Center for Quality of Life and Applied Psychology, Key Laboratory for Quality of Life and Psychological Assessment and Intervention, Guangdong Medical University, Dongguan 523808, China. [2]School of Nursing, The University of Hong Kong, Hong Kong, China. [3]Dongguan Enghth'Hospital, Dongguan 523325, China. [4]The First Affiliated Hospital of Kunming Medical University, Kunming 650100, China.

References

1. Dai P, Chang W, Xin Z, Cheng H, Ouyang W, Luo A. Retrospective Study on the Influencing Factors and Prediction of Hospitalization Expenses for Chronic Renal Failure in China Based on Random Forest and LASSO Regression. Front Public Health. 2021;9:678276. https://doi.org/10.3389/fpubh.2021.678276.
2. Ku do Y, Park YS, Chang HJ, Kim SR, Ryu JW, Kim WJ. Depression and life quality in chronic renal failure patients with polyneuropathy on hemodialysis. Ann Rehabil Med. 2012;36(5):702-707. https://doi.org/10.5535/arm.2012.36.5.702.
3. Chang JS, Chang EJ, Jung CH. The Factors Affecting General Coping Styles in Chronic Renal Failure Patients. J Korean Neuropsychiatr Assoc. 2003;42(3):396–402.
4. Yang Z, Huang XP, Weng M, Xu CZ, Chen Y, Pan HY, et al. Validity evaluation of the quality of life instruments for chronic diseases-chronic renal failure scale (QLICD-CRF). J Guangdong Med Coll. 2014;32(03):401–4.
5. Yang Z, Huang XP, Wang CH, Weng M, Li XM, Bai YK, et al. Reliability Evaluation of the Quality of Life Instrument for Patients with Chronic Renal Failure (QLICD-CRF) based on Classical Test Theory and Generalizability Theory. China Health Stat. 2015;32(02):224–6.
6. Brennan RL. Generalizability theory and classical test theory. Appl Meas Educ. 2011;24(1):1–21.
7. Lie DA, Richter-Lagha R, Forest CP, Walsh A, Lohenry K. When less is more: validating a brief scale to rate interprofessional team competencies. Med Educ Online. 2017;22(1):1314751. https://doi.org/10.1080/10872981.2017.1314751.
8. Sele P, Hoffart A, Bækkelund H, Øktedalen T. Psychometric properties of the International Trauma Questionnaire (ITQ) examined in a Norwegian trauma-exposed clinical sample. Eur J Psychotraumatol. 2020;11(1):1796187. https://doi.org/10.1080/20008198.2020.1796187.
9. Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. Educ Meas Issues Pract. 1993;12(3):38–47.
10. Yang ZM. Generalization theory of evaluation and its application. Educ Sci Press. 2003;203:18–219.
11. Sun KJ, Li PH, Zhou P. Classical versus generalizability theory of measurement. Examinations Res. 2007;3(04):50–63.
12. Lord FM. A theory of tests scores. Psychometric Monographs. 1952, No. 7.

13. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah: Lawrence Erlbaum Associates; 2000.
14. Li G, Pan Y, Wang W. Using Generalizability Theory and Many-Facet Rasch Model to Evaluate In-Basket Tests for Managerial Positions. Front Psychol. 2021;12:660553. https://doi.org/10.3389/fpsyg.2021.660553.
15. Vahedi S. World Health Organization Quality-of-Life Scale (WHOQOL-BREF): Analyses of Their Item Response Theory Properties Based on the Graded Responses Model. Iran J Psychiatry. 2010;5(4):140–53.
16. Ishihara M, Harel D, Levis B, Levis AW, Riehm KE, Saadat N, et al. Shortening self-report mental health symptom measures through optimal test assembly methods: Development and validation of the Patient Health Questionnaire-Depression-4. Depress Anxiety. 2019;36(1):82–92.
17. Levis AW, Harel D, Kwakkenbos L, Carrier ME, Mouthon L, Poiraudeau S, Bartlett SJ, Khanna D, Malcarne VL, Sauve M, van den Ende CH, Poole JL, Schouffoer AA, Welling J, Thombs BD; the Scleroderma Patient-Centered Intervention Network Investigators. Using Optimal Test Assembly Methods for Shortening Patient-Reported Outcome Measures: Development and Validation of the Cochin Hand Function Scale-6: A Scleroderma Patient-Centered Intervention Network Cohort Study. Arthritis Care Res (Hoboken). 2016;68(11):1704-1713.
18. Van der Linden WJ. Linear models for optimal test design. New York: Springer; 2006.
19. Harel D, Baron M. Methods for shortening patient-reported outcome measures. Stat Methods Med Res. 2019;28(10-11):2992–3011.
20. Wan CH. Evaluation and application of quality of life in patients with chronic diseases: Science Press; 2015.
21. Byrne BM. Structural equation modeling with Mplus: basic concepts, applications, and programming. London: Routledge Academic; 2012.
22. Hays RD, Hayashi T. Beyond internal consistency reliability: Rationale and use's guide for multi-trait analysis program on the microcomputer. Behav Res Methods Instrum Compu. 1990;22:167–75.
23. Nußbaum A. Multivariate generalizability theory in educational measurement: an empirical study. Appl Psychol Meas. 1984;8(2):219–30. https://doi.org/10.1177/014662168400800211.
24. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess. 2009;13:1–177.
25. Ware JE Jr, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol. 1998;51(11):945–52. https://doi.org/10.1016/s0895-4356(98)00085-7.
26. Hair JF, Jr, Black WC, Babin BJ. Multivariate data analysis (seventh edition) Edinburgh: Pearson; 2014.
27. Qi SQ, Dai HQ, Ding SL. Modern education and psychological measurement theory. Beijing: Beijing Higher Education Press; 2002.
28. Bollen KA. Structural equations with latent variables. New York: John Wiley; 1989.
29. Muthén L, Muthén B. Mplus User's Guide: Statistical Analysis With Latent Variables. Mplus User's Guide (Seventh Edition). 2012;1–850.
30. Brennan RL. Generalizability theory. New York: Springer-Verlag New York Inc; 2001.
31. Brennan RL. Manual for mGENOVA (Version 2.1). Iowa City: Iowa Testing Programs; 2001. p. 50.
32. Rizopoulos D. ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses. J Stat Software. 2006;17(5):1–25 Available online: http://www.jstatsoft.org/v17/i05/.
33. lp_solve, Kjell Konis and Florian Schwendinger . lpSolveAPI: R Interface to 'lp_solve'Version 5.5.2.0. 2020.R package version 5.5.2.0-17.7. Available online: https://CRAN.R-project.org/package=lpSolveAPI. Accessed 6 Nov 2022.
34. R Core Team. R: A language and environment for statistical computing. 2021. Available online: https://www.r-project.org/. Accessed 6 Nov 2022.
35. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24(3):69–71. https://doi.org/10.2166/wh.2012.000.
36. Cassidy MJD, Kerr DNS. The assessment of the patient with chronic renal insufficiency. In: Stewart C, Davidson Alex M, Grunfeld J-P, David K, Ritz E, editors. Oxford Textbook of Clinical Nephrology. Oxford: Oxford Medical Publications; 1992. p. 1155–60.
37. Baker FB. The Basics of Item Response Theory. 2nd ed. College Park: ERIC Clearinghouse on Assessment and Evaluation; 2001. p. 108–9.
38. Hu T, Wang B, Liao X, Wang S. Clinical features and risk factors of pruritus in patients with chronic renal failure. Exp Ther Med. 2019;18(2):964–71. https://doi.org/10.3892/etm.2019.7588.
39. Blood AD, Park YS, Lukas RV, Brorson JR. Neurology objective structured clinical examination reliability using generalizability theory. Neurology. 2015;85(18):1623–9. https://doi.org/10.1212/WNL.0000000000002053.
40. Sekely A, Taylor GJ, Bagby RM. Developing a short version of the Toronto Structured Interview for Alexithymia using item response theory. Psychiatry Res. 2018;266:218–27. https://doi.org/10.1016/j.psychres.2018.03.002.
41. Sahin A, Anil D. The effects of test length and sample size on item parameters in item response theory. Educ Sci. Theory Pract. 2017;17:321–35. https://doi.org/10.12738/estp.2017.1.0270.

Publisher's Note