Qual Life Res (2018) 27:1041–1054
DOI 10.1007/s11136-017-1553-1

Modernizing quality of life assessment: development of a multidimensional computerized adaptive questionnaire for patients with schizophrenia

Pierre Michel[1,2] · Karine Baumstarck[1] · Christophe Lancon[1] · Badih Ghattas[1,2] · Anderson Loundou[1] · Pascal Auquier[1] · Laurent Boyer[1]

Accepted: 14 March 2017 / Published online: 25 March 2017

Abstract
Objective Quality of life (QoL) is still assessed using paper-based and fixed-length questionnaires, which is one reason why QoL measurements have not been routinely implemented in clinical practice. Providing new QoL measures that combine computer technology with modern measurement theory may enhance their clinical use. The aim of this study was to develop a QoL multidimensional computerized adaptive test (MCAT), the SQoL-MCAT, from the fixed-length SQoL questionnaire for patients with schizophrenia.
Methods In this multicentre cross-sectional study, we collected sociodemographic information, clinical characteristics (i.e., duration of illness, the PANSS, and the Calgary Depression Scale), and quality of life (i.e., SQoL). The development of the SQoL-CAT was divided into three stages: (1) multidimensional item response theory (MIRT) analysis, (2) multidimensional computerized adaptive test (MCAT) simulations with analyses of accuracy and precision, and (3) external validity.
Results Five hundred and seventeen patients participated in this study. The MIRT analysis found that all items displayed good fit with the multidimensional graded response model, with satisfactory reliability for each dimension. The SQoL-MCAT was 39% shorter than the fixed-length SQoL questionnaire and had satisfactory accuracy (levels of correlation >0.9) and precision (standard error of measurement <0.55 and root mean square error <0.3). External validity was confirmed via correlations between the SQoL-MCAT dimension scores and symptomatology scores.
Conclusion The SQoL-MCAT is the first computerized adaptive QoL questionnaire for patients with schizophrenia. Tailored for patient characteristics and significantly shorter than the paper-based version, the SQoL-MCAT may improve the feasibility of assessing QoL in clinical practice.

Keywords Schizophrenia · Quality of life · Patient-reported outcome · Multidimensional computerized adaptive testing · Item response theory · Psychometrics

Introduction

Quality of life (QoL) measurements are considered important in the evaluation of disease progression, treatment options, and the management of care provided to patients with schizophrenia [1–3]. It is recognized that reducing symptoms does not indicate successful management of all of the facets that patients consider to be important in their life, and QoL may add complementary information to traditional clinical assessments, which may not adequately reflect patients' perceptions [3]. QoL has been reported to be an independent predictor for long-term symptomatic remission, functional recovery, and disability [4, 5].

Electronic supplementary material The online version of this article (doi:10.1007/s11136-017-1553-1) contains supplementary material, which is available to authorized users.

✉ Pierre Michel
pierre.michel@univ-amu.fr

[1] Aix-Marseille University, EA 3279 – Public Health, Chronic Diseases and Quality of Life - Research Unit, 13005 Marseille, France

[2] Aix-Marseille University – I2M UMR 7373 – Mathematics Institute of Marseille, 13009 Marseille, France

Despite the acknowledged need to consider QoL issues in patients with schizophrenia, QoL measurement has not been routinely implemented in clinical practice [6]. Of the various reasons for this lack of implementation [3], one concern expressed by clinicians was the administrative burden associated with the general lack of efficiency in evaluating QoL within the workplace [5, 7, 8]. Most QoL questionnaires are still paper-based, making it challenging for professionals to obtain QoL scores in an efficient real-time manner. These questionnaires are often too lengthy and fixed in content (i.e., asking the same questions to all patients regardless of their health characteristics), leading to a high survey burden for the patients and to substantial problems with missing data [4].

The provision of new QoL measures that combine computer technology with modern measurement theory could reduce the administrative obstacles and lessen the survey burden for patients. Several experiments were carried out in recent years, the most notable is the one funded by the National Institutes of Health since 2004 called the Patient-Reported Outcomes Measurement Information System (PROMIS) [9, 10]. Methods based on multidimensional item response theory (MIRT) models and multidimensional computerized adaptive testing (MCAT) can overcome the problems posed by paper-based and fixed-length QoL questionnaires [11, 12]. MCAT allows for the administration of only the items that offer the most relevance to a given patient, reducing the length of the questionnaire and the completion time in addition to maintaining the test's precision [13–15]. MCAT has been recently applied to measure health problems (e.g., symptomatology, fatigue, physical and emotional functioning) in various chronic diseases (e.g., child mental health, cancer) [16–18] but is not available for measuring QoL in patients with schizophrenia.

The aim of this study was to develop MCAT for patients with schizophrenia from a fixed-length available QoL questionnaire. Our study focused on the Schizophrenia Quality of Life Questionnaire (SQoL) [19], which is a widely used QoL questionnaire for schizophrenia with application in various populations and settings (e.g., patients with severe cognitive impairment [20, 21], homeless patients [22], inpatients [23], and neuroimaging studies [24, 25]). Compared to other QoL questionnaires, this instrument has three important characteristics: (a) it is anchored in an explicit conceptual approach, the Calman's approach, defining QoL as the discrepancy between expectations and the current life experience [26]; (b) it specifically reflects the perspectives of patients with schizophrenia, with items generated by individual interviews with patients [27]; and (c) it is available in multiple languages [28, 29].

Methods

Questionnaire

The SQoL is a specific, self-administered and multidimensional QoL questionnaire designed for people with schizophrenia. It includes 41 items describing 8 dimensions: Psychological Well-Being (PsW, 10 items), Self-Esteem (SE, 6 items), Family Relationships (RFa, 5 items), Relationships with Friends (RFr, 5 items), Resilience (RE, 5 items), Physical Well-Being (PhW, 4 items), Autonomy (AU, 4 items) and Sentimental Life (SL, 2 items). It also includes a total score (Index), computed as the mean of the 8 SQoL dimensions scores. The 8 dimensions and the Index score range from 0 to 100; higher scores indicate better QoL [19]. The text corresponding to each item is presented in Additional File 2.

Study design, setting, and population

We established a database from four studies conducted by members of the SQoL Group (i.e., the developers of the SQoL, including public health professionals, psychiatrists, psychologists, and statisticians) in which the SQoL was used to assess patients' QoL. The database included a total of 517 in- and outpatients recruited from four psychiatric hospitals in France: Lyon (one hospital), Marseille (two), and Toulon (one). The inclusion criteria were a diagnosis of schizophrenia according to the Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV-TR) criteria [30], an age over 18 years, the provision of informed consent to participate in the studies, and French as one's native language. The exclusion criteria were a diagnosis other than schizophrenia on Axis I of the DSM-IV, decompensated organic disease, and mental retardation. These projects were conducted in accordance with the Declaration of Helsinki and French Good Clinical Practices [31, 32].

Data collection

In addition to the SQoL questionnaire, the following data were collected:

a. Sociodemographic information: age (years), gender (male, female), educational level (less than 12 years, greater than 12 years), and patient status (inpatient, outpatient).
b. Clinical data: disease duration (years); severity of psychotic symptoms based on the Positive and Negative Syndrome Scale (PANSS), which comprises the following three factors: positive, negative, and general psychopathology (PANSS P, PANSS N, and PANSS G, respectively) [33, 34]; and depression, assessed

using the Calgary Depression Scale for Schizophrenia (CDSS) [35].

Development of the SQoL-MCAT

The development of the SQoL-CAT was divided into three stages: (1) MIRT analysis, (2) MCAT simulations with analyses of accuracy and precision, and (3) external validity of the SQoL-MCAT. A detailed description of the statistical methods can be found in a previous publication [36].

MIRT analysis

All the items of the SQoL were considered as effect or reflective indicators and were thus allowed to be used in a latent trait model [37]. Following the work of Costa [38], we may assume that the items of the SQoL are the manifestations of underlying construct. The removal of items or the addition of homogeneous items does not change the construct of interest. The multidimensional construct was defined independently of the items, and the items were designed to represent this definition by the patients themselves. The 8-factor structure was validated using a confirmatory factor analysis (CFA) model with weighted least squares estimation (i.e., adapted to ordinal data) using the MPlus software [39], to assess construct validity. Model fit was assessed using the root mean square error of approximation (RMSEA <0.05 expected) and the comparative fit index (CFI >0.9 expected) [40]. We calibrated a multidimensional graded response model [41, 42] (MGRM) to our data, which is a "between-item" multidimensional model (i.e., each item loads on only one dimension). According to Adam's work [43], a questionnaire is considered multidimensional between items if each item belongs to precisely one dimension. In addition, the latent variables measured by the different subscales are assumed to correlate. The SQoL-41 contains 8 dimensions that measure related but distinct latent dimensions (e.g., physical well-being, psychological well-being, sentimental life, family relationship…). Importantly, the MGRM is appropriate to use when item responses can be characterized as ordered categorical responses such as in the SQoL-41 [44]. In this model, the multiple latent traits are allowed to correlate. Item parameters were estimated using the Metropolis–Hastings Robbins–Monro (MH-RM) [45] method, as implemented in the R package mirt [46]. The MH-RM algorithm allowed us to use stochastically imputed complete-data likelihood with an assumed population multivariate normal distribution, which is used to produce latent trait estimates and the imputed data. The item parameters were estimated using the complete-data log-likelihood function. Item fit was assessed using the $S-\chi^2$statistic [47], which is adapted to multidimensional items. Items having a p value <0.05 were considered to show misfit. Latent trait scores were estimated by Bayesian maximum a posteriori (MAP) estimation [11]. Item and test information were calculated. The percentage of the item information was computed as the height at the sample mean divided by the height of the test information function at the sample mean. We also assessed for each dimension empirical marginal reliability estimates [48]: coefficients greater than or equal to 0.7 were considered satisfactory [49].

The unidimensionality of each dimension was assessed using a Rasch analysis. The goodness-of-fit statistics (inlier-sensitive fit, INFIT) ensured that all items of the scale measured the same concept and were productive for measurement (expected values between 0.5 and 1.5) [50]. Monotonicity was checked using visual analysis of item characteristic curves. Mokken scale analysis was performed to compute scalability coefficients for each item and each dimension. As stated mathematically in previous works [51, 52], all scalability coefficients range between 0 and 1 (with 1 indicating a perfect scalability). Since these coefficients capture basically the number of violations of Guttman rules, some rule of thumb is usually applied [53]. Values greater than 0.4 were expected for each dimension, while values greater than 0.3 were expected for each item.

Differential item functioning (DIF) was explored to identify systematic errors for the 41 SQoL items due to group biases, as reported in previous studies (e.g., paranoid symptoms, lack of insight) [54, 55]: gender (male versus female), presence of paranoid symptoms (paranoid schizophrenia versus others), and levels of insight (using the SUMD scale). DIF was assessed using an IRT-based iterative ordinal logistic regression model, as implemented in the R package lordif [56]. The LR χ^2ratio test was used to detect an overall DIF effect at the level $\alpha = 0.05$. In cases of statistical significance, Zumbo's DIF classification was used to assess the DIF magnitude by computing ΔR^2. The DIF magnitude was considered negligible if $\Delta R^2 < 0.13$, moderate if $0.13 < \Delta R^2 \leq 0.26$ and large if $\Delta R^2 > 0.26$ [57].

MCAT simulations with analyses of accuracy and precision

We implemented a real-data simulation approach, i.e., complete response patterns to the 41 items of the SQoL were used to simulate the conditions of the MCAT assessment. We used the responses contained in the item bank to simulate the adaptive administration of items, as in a previous work [58]. A sensitivity analysis to test the robustness of our findings was performed after multiple imputation under the assumption that missing data were missing at random [59, 60]. Five imputations were performed in our initial data sample. The analysis used the imputed datasets with the estimated model parameters and the latent

trait estimates to impute plausible missing data values. The algorithm of the MCAT was based on Kullback–Leibler information item selection [61]. For the starting item, we used the item with the highest percentage of test information at the mean of the patient sample. Item selection depended on the responses to the previous items in the questionnaire, which were taken from the observed data. After each item was completed, the latent trait scores were estimated using Bayesian MAP estimation. The stopping rule employed was the pre-specified level of measurement precision using the standard error of measurement (SEM). An acceptable range has been defined as 0.33 to 0.55, corresponding to reliability coefficients between 0.70 and 0.90 [57].

The MCAT was run under three levels of minimally required SEM (i.e., minimum, middle, and maximum values: 0.33, 0.44, and 0.55, respectively), i.e., the MCAT procedure was stopped when the SEM threshold was reached for each dimension. For these 3 simulations, MCAT scores were calculated, and accuracy and precision were then assessed. Accuracy was assessed using the level of correlation between the MCAT scores and the latent trait scores based on the full set of items (levels of correlation >0.9 were expected for each dimension). Precision was assessed using the root mean square error (RMSE). The RMSE is the square root of the mean square of all the errors. The error is the gap between the latent traits estimated by the MCAT and the latent traits estimated by the full item bank. The formula is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2},$$

where y_i is the full item bank latent trait standardized estimate and $\hat{y}_i$ is the MCAT latent trait estimate. Both scores are on the same metric, such as a CAT score, which is on the same metric as the item bank on which it is based. Smaller values of RMSE represent better measurement precision, and RMSE values lower than or equal to 0.3 indicate excellent measurement precision [64]. The final SQoL-MCAT was selected considering the lowest number of items that matched with the most satisfactory level of accuracy and precision.

External validity of the SQoL-MCAT

The divergent validity was determined by exploring the relationships between the SQoL-MCAT scores and sociodemographic (i.e., age, gender, educational level, patient status) and clinical (i.e., disease duration, PANSS scores, and CDSS score) characteristics using *t* tests and Pearson's correlations. In accordance with previous studies [19, 20], we expected the following findings: the SQoL-MCAT dimension scores (1) should not differ with sociodemographic characteristics; (2) should be higher in outpatients than in inpatients; (3) should not depend on the duration of the disease; and (4) should be negatively correlated with the severity of the disease.

All statistical analyses were performed using R [65].

Results

The study sample included 517 patients with schizophrenia. The mean age was 36.5 years (Standard Deviation = 10.8), and 29.8% ($n = 154$) were female. The mean duration of illness was 13.8 years (SD = 9.3), and the patients had a moderate severity of psychotic symptoms, with a total PANSS score of 69.6 (SD = 18.4). All of these characteristics are shown in Table 1.

MIRT analysis

The eight-factor structure tested in the CFA model showed satisfactory fit indices (RMSE = 0.05, CFI = 0.95). The item characteristics (i.e., missing values,

Table 1 Descriptive characteristics of the study sample

Sociodemographic data	
Age ($M \pm$ SD)	
Years	36.5 ± 10.9
Gender (N [%])	
Male	362 [70%]
Female	154 [30%]
Educational level (N [%])	
<12 years	162 [79%]
≥12 years	43 [21%]
Patient status (N [%])	
Inpatient	131 [64%]
Outpatient	74 [36%]
Clinical data	
Disease duration	
Years	13.8 ± 9.3
PANSS ($M \pm$ SD)	
Total score	69.6 ± 18.4
Positive score	15.7 ± 6.1
Negative score	19.2 ± 7.0
General score	35.8 ± 9.7
CDSS ($M \pm$ SD)	
Total score	3.1 ± 3.6

PANSS Positive and Negative Syndrome Scale, *CDSS* Calgary Depression Scale score for Schizophrenia

Table 2 Item characteristics

Item	Dimension	Missing values (%)	*a*	*b*1	*b*2	*b*3	*b*4	Information* (%)	INFIT	$S-\chi^2 p$ value
1	SE	2.71	2.136	−1.64	−0.92	−0.13	0.92	2.39	0.87	0.20
2	RE	4.26	1.731	−1.65	−0.84	−0.30	0.62	1.62	0.84	0.03
3	RE	4.84	1.58	−1.41	−0.71	−0.18	0.71	1.39	0.84	0.91
4	RE	5.22	1.814	−1.19	−0.44	0.31	1.26	1.82	0.85	0.85
5	SE	1.93	2.2	−1.65	−0.89	−0.04	1.02	2.53	0.78	< 0.01
6	SE	2.71	2.367	−1.67	−0.92	−0.03	1.15	2.87	0.76	0.09
7	SE	2.13	2.826	−1.92	−1.01	−0.12	1.10	3.89	0.69	0.24
8	SE	3.87	2.654	−1.57	−0.70	0.28	1.44	3.55	0.90	0.02
9	AU	2.90	3.045	−2.15	−1.21	−0.42	1.35	4.22	0.61	0.22
10	AU	2.32	3.376	−2.54	−1.56	−0.56	1.50	4.82	0.63	0.32
11	RE	4.06	1.988	−1.10	−0.37	0.41	1.41	2.15	0.87	0.30
12	RE	8.51	1.609	−1.41	−0.77	−0.13	0.84	1.44	0.84	0.67
13	**AU**	**3.09**	**1.137**	**−1.08**	**−0.56**	**−0.16**	**0.78**	**0.74**	**1.00**	**0.89**
14	SL	6.19	2.314	−1.04	−0.28	0.43	1.47	2.83	0.60	0.53
15	PhW	2.13	2.319	−1.79	−1.07	−0.06	1.27	2.73	0.68	0.07
16	PhW	4.26	2.553	−1.54	−0.72	0.21	1.44	3.32	0.68	0.15
17	PhW	6.19	1.254	−0.68	−0.10	0.36	1.08	0.90	0.96	0.65
18	PhW	3.48	1.832	−1.39	−0.83	−0.08	1.25	1.79	0.90	0.04
19	Rfa	4.45	2.151	−1.58	−0.99	−0.43	0.91	2.33	1.00	0.22
20	Rfa	6.00	2.9	−1.92	−1.30	−0.65	0.91	3.78	0.65	0.06
21	Rfa	7.35	3.136	−1.96	−1.14	−0.31	1.27	4.53	0.71	0.27
22	AU	5.61	1.177	−1.07	−0.62	−0.09	0.77	0.80	0.98	0.25
23	Rfa	4.06	1.706	−1.39	−0.97	−0.52	0.76	1.50	1.11	0.12
24	**Rfa**	**6.00**	**3.378**	**−2.01**	**−1.30**	**−0.49**	**1.28**	**4.92**	**0.64**	**0.26**
25	RFr	4.64	2.12	−0.96	−0.33	0.25	1.36	2.37	0.88	0.22
26	RFr	2.90	1.535	−1.18	−0.68	−0.21	0.97	1.29	1.10	0.17
27	RFr	7.54	2.775	−1.43	−0.49	0.06	1.65	3.72	0.67	0.17
28	RFr	7.35	2.588	−1.56	−0.74	−0.04	1.65	3.28	0.80	0.60
29	RFr	6.77	2.575	−1.23	−0.47	0.10	1.50	3.30	0.79	0.12
30	SL	7.54	2.314	−0.64	−0.01	0.67	1.53	2.71	0.57	0.94
31	SE	2.71	1.506	−0.84	−0.35	0.16	1.17	1.26	1.23	0.82
32	PsW	3.29	1.841	−1.13	−0.56	0.07	0.62	1.82	0.93	0.14
33	PsW	5.61	2.359	−1.68	−1.02	−0.47	0.12	2.54	0.81	0.23
34	PsW	5.03	2.109	−1.63	−0.91	−0.41	0.23	2.16	0.92	0.23
35	PsW	5.03	1.615	−1.10	−0.57	−0.16	0.48	1.41	1.06	0.85
36	PsW	3.87	1.841	−1.32	−0.59	−0.03	0.52	1.80	0.91	0.30
37	PsW	4.84	1.8	−1.39	−0.76	−0.27	0.33	1.68	0.98	0.36
38	PsW	4.64	2.11	−1.46	−0.85	−0.35	0.20	2.14	0.87	0.76
39	PsW	5.80	1.442	−1.08	−0.59	−0.18	0.35	1.13	1.17	0.19
40	PsW	3.87	1.983	−1.43	−0.76	−0.25	0.47	2.03	0.88	0.88
41	PsW	4.45	2.268	−1.66	−1.02	−0.42	0.28	2.47	0.87	0.46

SE self-esteem, *RE resilience, AU* autonomy, *SL* sentimental life, *PhW* physical well-being, *RFa* relationships with family, *RFr* relationships with friends, *PsW* psychological well-being, *a* item discrimination parameter associated with the corresponding dimension of the item, *b1, b2, b3, b4* item difficulty threshold parameters

The most and the least informative items are in bold

*Percentage of the total information provided by the item

item parameters, item information, unidimensional and multidimensional item fit) are presented in Table 2, and the latent trait score distributions for each dimension are presented in Fig. 1. The correlation matrix of the eight-dimensional latent distributions is shown in Additional File 1. Item 24 from the Family relationships dimension ("My family pays attention to me") provided the highest amount of information, and item 13 from the Autonomy dimension ("I'm able to go out (cinema, walking, restaurant…)") provided the least amount of information. The coefficients based on empirical marginal reliability estimates were high for all dimensions (ranging from 0.80 to 0.92).

Item fit was satisfactory for both the unidimensionality of each dimension (INFIT values ranged from 0.57 to 1.23 for each item) and for global multidimensionality (only 3 items had p values < 0.05). Misfitting items according to the $S - \chi^2$ statistic were not rejected because their proportion of information was high or their INFIT was within the acceptance range, indicating that they should be important during the MCAT assessment. Results of Mokken analysis were satisfactory. All scalability coefficients related to dimensions were greater than 0.4, indicating that all items discriminated well among different values of latent traits. The scalability coefficient of each item was greater than 0.3, indicating that each item was coherent in its dimension.

Of the 123 tests performed (41 items by three confounding factors), six exhibited overall DIF (see Appendix). Following Zumbo's DIF classification, no items were flagged for moderate or large DIF magnitudes. Some items were flagged for negligible DIF magnitudes: three for gender (items 18, 24, and 33), one for the presence of paranoid symptoms (item 27), and two for level of insight (items 3 and 26). Given the negligible DIF magnitudes, all 41 items were assumed to keep their invariance according to these characteristics. DIF results are provided in Appendix.

MCAT simulations with analyses of accuracy and precision

Real-data simulations were performed on 348 patients with complete response patterns to the 41 items of the SQoL. The accuracy and precision indicators for each simulation are presented in Table 3.

For each simulation, correlations between the MCAT dimension scores and the scores based on the full set of items were considered. All eight dimensions had satisfactory accuracy, with correlations higher than 0.90. Even for the less restrictive model (i.e., SEM < 0.55), the accuracy was satisfactory, with correlations higher than 0.94.

The precision improved when the MCAT simulations tested a lower threshold of SEM. However, the RMSE values were satisfactory for the three simulations, with RMSE < 0.3 (except for the RE dimension in the model based on SEM < 0.55, in which the RMSE was slightly higher than 0.3).

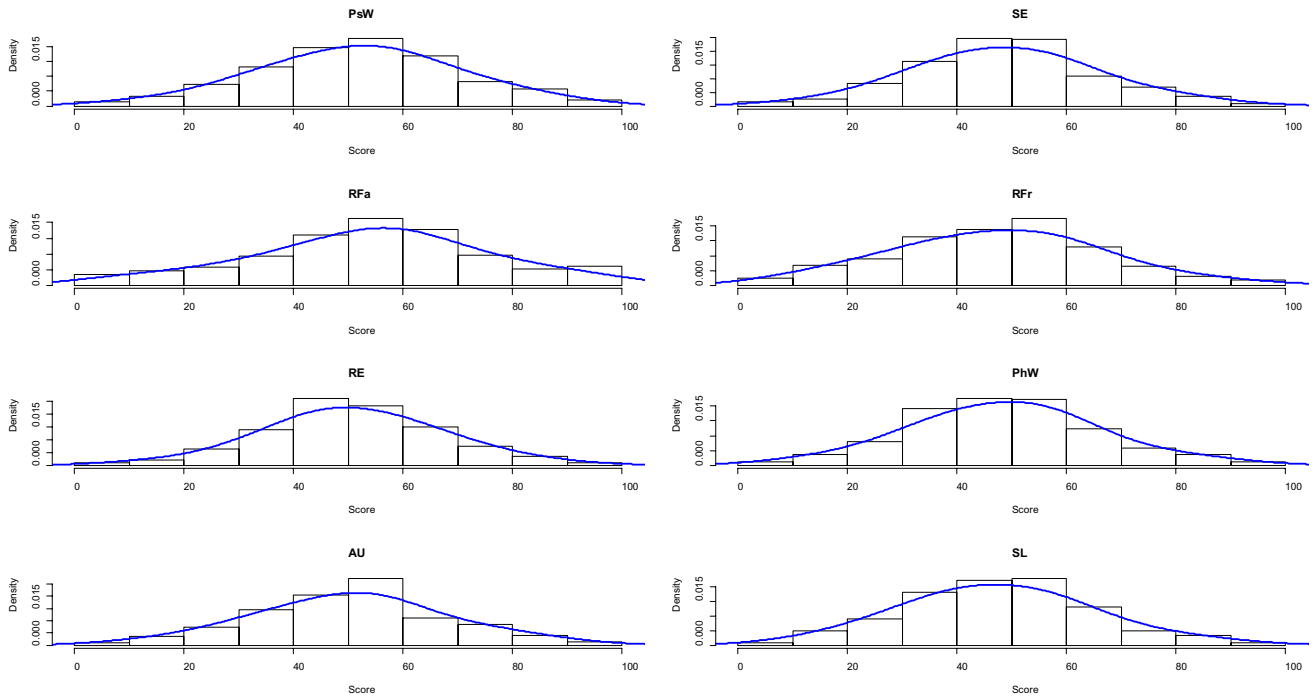

Fig. 1 IRT score distribution for each SQoL dimension

Table 3 Mean score, accuracy and precision indicators, and mean number of items for each MCAT simulation

Precision level	Indicator	PsW	SE	Rfa	RFr	RE	PhW	AU	SL
SEM < 0.33	Mean score	51.56	48.66	53.45	47.56	50.68	48.37	51.01	46.94
	Accuracy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	RMSE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	SEM	0.30	0.28	0.32	0.34	0.37	0.37	0.37	0.43
	Mean number of items	10.0	6.0	5.0	5.0	5.0	4.0	4.0	2.0
SEM < 0.44	Mean score	51.45	48.73	53.52	47.63	50.87	48.41	51.12	46.96
	Accuracy	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00
	RMSE	0.12	0.05	0.05	0.06	0.15	0.09	0.09	0.02
	SEM	0.32	0.29	0.33	0.35	0.40	0.38	0.38	0.43
	Mean number of items	8.3	5.5	4.6	4.5	3.8	3.2	2.9	2.0
SEM < 0.55	Mean score	51.72	48.68	53.59	47.53	50.84	48.39	51.05	46.92
	Accuracy	0.95	1.00	0.99	0.99	0.94	0.98	0.99	0.99
	RMSE	0.27	0.09	0.13	0.12	0.32	0.17	0.13	0.10
	SEM	0.39	0.30	0.34	0.36	0.48	0.40	0.39	0.44
	Mean number of items	4.2	5.1	3.8	3.9	1.6	2.3	2.1	1.9

PsW psychological well-being, *SE* self-esteem, *RFa* relationships with family, *RFr* relationships with friends, *RE* resilience, *PhW* physical well-being, *AU* autonomy, *SL* sentimental life

The average number of items was 25 (SD = 5) for the model based on SEM <0.55, 35 (SD = 6) for the model based on SEM <0.44, and 41 (SD = 0) for the model based on SEM <0.33. For the latter, all 41 items were asked, for each patient. For each SEM criterion, all patients had at least one item from each dimension administered.

The model based on a level of precision of SEM <0.55 was defined as the most satisfactory MCAT simulation because this model was associated with the lowest number of items and maintained satisfactory levels of accuracy and precision. Item exposure for this simulation is presented in Fig. 2. All items were administered at least once, while eighteen were administered more than 9 times out of 10 (items 1, 5–11, 14–16, 20, 21, 24, 27–29, 41).

The sensitivity analysis after multiple imputations on the 517 patients presented findings similar to those obtained from the complete response patterns.

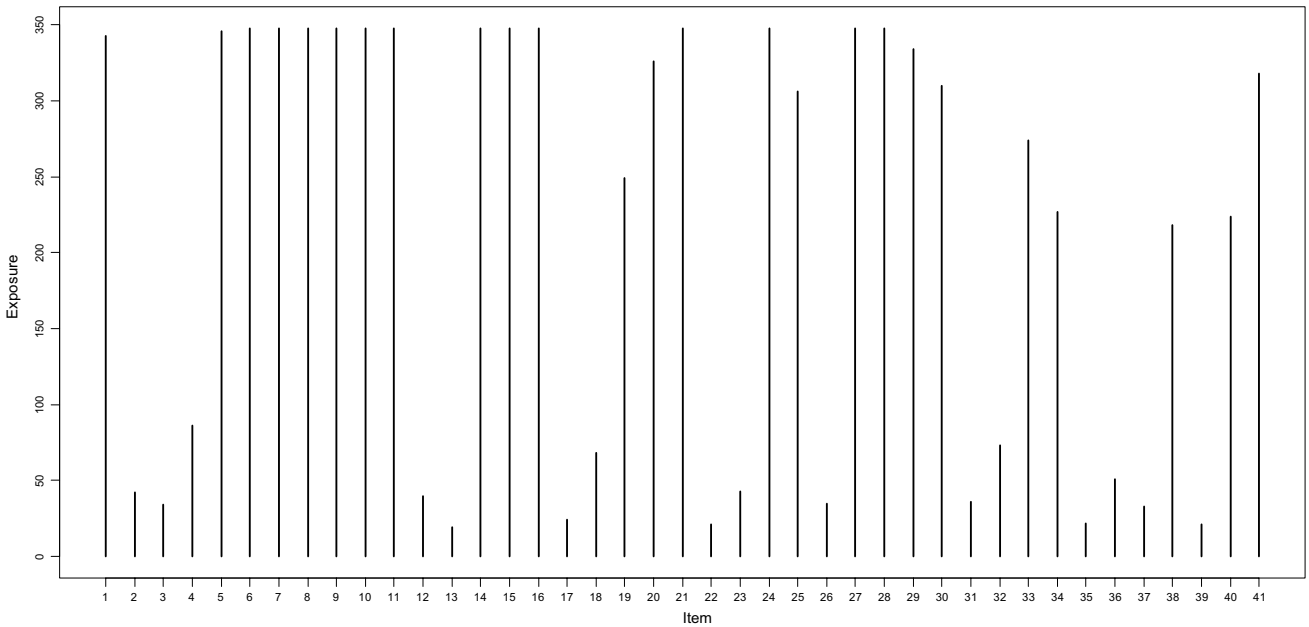

Fig. 2 Item exposure of the precision-based SQoL-MCAT

External validity of the SQoL-MCAT

Convergent and divergent validity were assessed for the precision-based SQoL-MCAT (SEM < 0.55), and the results are shown in Table 4.

The SQoL-MCAT dimension scores were not correlated with the duration of illness. Concerning the correlations between SQoL-MCAT dimension scores and clinical features (PANSS and CDSS), higher QoL levels were globally associated with lower levels of severity.

Discussion

QoL measures provide clinicians with information regarding the general health of patients who might otherwise go unrecognized, but they have not been routinely implemented in psychiatric practice. The logistics of obtaining QoL data are largely less developed than those of other clinical indicators (e.g., biology, imaging). However, obtaining QoL data in an efficient real-time manner could actually be easy with the use of new technology [7]. Adaptive testing associated with computer technology may enhance the use of QoL measures in clinical decision making. In contrast to traditional fixed-length QoL assessments, MCAT selects items that are most relevant for a patient, and the computer scores the responses in real time on a standardized metric that permits comparisons among patients answering different questions from the same pool of items. The SQoL-MCAT is the first multidimensional computerized adaptive QoL questionnaire specific to patients with schizophrenia, and it asks fewer questions and requires substantially less time to complete. Logistic barriers should now be considered to implement these new QoL instruments in clinical settings. Indeed, obtaining QoL data in an efficient real-time manner implies computer stations and hand-held devices that are not systematically available in setting marked by limited resources [3, 8].

The SQoL-MCAT had satisfactory precision and accuracy. All of the SQoL-MCAT dimensions had levels of correlation that were higher than 0.9 with the IRT dimension scores based on the full set of items and had RMSEs lower than 0.3, except for in one dimension (i.e., RE). However, the RMSE for the RE dimension was acceptable.

Table 4 External validity of the SQoL-MCAT

	PsW	SE	RFa	RFr	RE	PhW	AU	SL
Age	0.045	0.067	**−0.190****	**−0.118***	0.019	−0.011	0.031	0.024
Gender								
Male	51.75 ± 16.78	48.68 ± 16.44	54.42 ± 20.71	48.32 ± 18.41	50.75 ± 15.19	48.69 ± 16.90	51.20 ± 17.05	45.92 ± 16.36
Female	51.85 ± 17.73	48.82 ± 16.77	51.29 ± 23.36	45.67 ± 20.32	51.24 ± 15.13	47.77 ± 17.11	50.94 ± 18.17	49.18 ± 17.44
p value	0.956	0.943	0.242	0.256	0.782	0.647	0.900	0.108
Educational level								
<12 years	50.28 ± 18.22	47.85 ± 16.00	50.12 ± 22.75	45.05 ± 19.50	49.69 ± 14.22	46.66 ± 16.51	48.22 ± 18.14	46.40 ± 16.95
≥12 years	57.83 ± 12.57	53.78 ± 15.34	55.90 ± 16.69	51.72 ± 16.41	54.04 ± 15.48	54.76 ± 13.53	53.58 ± 18.89	51.39 ± 15.64
p value	**0.024**	0.110	0.178	0.102	0.230	**0.019**	0.229	0.188
Patient status								
Inpatient	48.6 ± 17.23	47.17 ± 15.36	50.86 ± 23.01	43.74 ± 20.3	48.76 ± 14.54	46.99 ± 16.43	45.79 ± 18.41	45.75 ± 16.61
Outpatient	57.51 ± 16.53	52.38 ± 16.71	52.04 ± 19.47	51.12 ± 15.71	53.78 ± 14.08	50.61 ± 15.81	55.56 ± 16.64	50.36 ± 16.78
p value	**0.008**	0.09	0.782	**0.047**	0.076	0.256	**0.006**	0.159
Disease duration	0.043	0.092	−0.061	−0.067	0.034	0.009	0.000	0.052
p value	0.592	0.248	0.443	0.399	0.670	0.907	0.997	0.510
PANSS								
Total score	−0.082	−0.075	−0.109	−0.066	−0.100	−0.101	−0.097	−0.089
Positive score	**−0.157****	**−0.171****	−0.062	**−0.177****	**−0.178****	−0.113	**−0.138***	**−0.123***
Negative score	**−0.239****	**−0.225****	**−0.170****	**−0.129***	**−0.232****	**−0.192****	**−0.189****	**−0.213****
General score	**−0.245****	**−0.225****	**−0.135***	**−0.181****	**−0.236****	**−0.187****	**−0.184****	**−0.219****
CDSS	**−0.311****	**−0.380****	−0.139	**−0.170***	**−0.338****	**−0.295****	**−0.216****	**−0.343****

PANSS Positive and Negative Symptom Scale, *CDSS* Calgary Depression Scale for Schizophrenia

In bold: statistically significant

* *p* value <0.05

** *p* value <0.01

Additionally, the external validity of the SQoL-MCAT was consistent with the external validity of the fixed-length SQoL [19, 20]. No significant differences were reported according to gender, which is consistent with previous studies [19, 20, 66]. As expected, comparisons by educational level and age were rarely significant. A higher educational level was associated with higher QoL levels in two dimensions (PsW and PhW). The results of surveys that have been conducted on this subject are often contradictory. Ruggeri et al. [67] showed that a low level of education was associated with a low level of QoL, while Reine et al. [68] showed that more educated patients reported lower QoL. Older age was significantly associated with worse scores on the social dimensions (RFa and RFr). According to Kemmler et al. [69], social problems, isolation, and even stigmatization of patients with schizophrenia tend to increase with age, while Folsom et al. [70] showed that older age was associated with greater mental QoL. The SQoL-MCAT scores were moderately correlated with symptomatology, suggesting that the SQoL-MCAT may add complementary information to traditional clinical measures.

In our study, we used the Bayesian MAP method to estimate the latent trait level for the initial estimation of IRT scores, for updating the scores during the CAT procedure and for the final estimation of CAT scores. There are two main approaches for latent trait estimation: maximum likelihood (ML) estimation and Bayesian estimation including MAP and expected a posteriori (EAP). Although choosing the MAP estimation method might be questionable and may lead to some bias [71], a previous study [72] showed that MAP yields better precision than ML and performs similarly or better than EAP. Moreover, according to recent findings [46], using EAP scores for models with more than three factors is generally not recommended because it results in slower estimation and less precision. We decided to use MAP scores instead of EAP scores given the high-dimensional SQoL structure.

Finally, this study provided a broader reflection on the development strategy of new QoL measurements based on CAT technology. CAT has proven to be efficient compared with fixed-length questionnaire measurements, including increases in precision and omission of non-informative questions. An important foundation of CAT is the development of unidimensional item banks that contain a large number of items covering the entire scope of a latent trait (e.g., fatigue, pain) [58, 73]. The construction of a QoL item bank is an important step in proposing QoL CAT. However, a QoL item bank requires substantial resources and time because several issues have remained unresolved: Is it possible to associate several QoL questionnaires that are based on different theoretical and conceptual backgrounds with the same bank? Can we associate generic and specific questionnaires? Should we associate questionnaires developed from the perspective of the patient with those from experts? Additionally, the multidimensional nature of QoL requires the development of all of the unidimensional attributes of QoL that should be calibrated; only then would the development of a multidimensional measure be possible. All of these issues need to be resolved, and they therefore present a delay in the development of a large QoL item bank and thus a multidimensional QoL CAT that could be based on such a bank. Pending completion of this major work, although the number of items is relatively small in QoL questionnaires compared with the number in item banks, the development of MCAT from available QoL questionnaires can be an attractive option given the financial and time resources. A small number of items is not as restrictive as in unidimensional IRT in MIRT analysis, following the work of Zhang [74].

Limitations and perspectives

In our study, the sample size was not sufficiently important to split into two subsamples. Our approach may suffer from overfitting and future studies should test our instrument in a new sample of patients. However, we performed a sensitivity analysis using a cross-validation procedure based on a learning sample containing two-thirds of the observations and a test sample with the remaining observations. The learning sample was used for model estimation and the test sample was used for CAT simulations. This procedure was repeated 100 times, and the results of the simulations were averaged. Overall, we obtained similar results, and the precision indicators remained stable. Despite the large overall sample size in this study, the representativeness of our sample should be discussed. The patients were mostly middle-aged males with mild disease severity and more than 5 years of illness duration. Our findings about psychometric properties may thus not be generalizable to all patients with schizophrenia, particularly the most severe patients (e.g., severe cognitive impairment). The extent to which the SQoL-MCAT remains relevant and valid for all patients with schizophrenia is a crucial issue. However, recent works have explored the validity of the SQoL in various populations and settings (i.e., severe cognitive impairment, homeless), confirming its satisfactory properties [21, 22, 75, 76]. These studies should be replicated using the SQoL-MCAT. Future research with different sample characteristics could improve the generalizability and applicability of the SQoL-MCAT.

In our approach, latent traits were assumed to be normally distributed. This choice is highly questionable in a clinical population. Recent research suggests that some latent traits may have different non-normal distributions

[77]. This issue could lead to new alternative simulations in future works using clinicians' expertise.

In our simulations, three different levels of measurement precision (i.e., SEM of 0.33, 0.44, and 0.55, which come down to a marginal reliability of 0.90, 0.81, and 0.70, respectively) were considered. The precision-based threshold "SEM < 0.33" was too restrictive because not all latent trait estimates reached this threshold, and the mean number of items administered was the same as that of the fixed SQoL-41 questionnaire. This issue could be controlled by adding constraints less restrictive for dimensions that cannot reach the precision criterion, by using a mixed SEM criterion (with a specific threshold for each dimension) or by adding another precision-based criterion, based on the deviation of the latent trait estimate observed after each item is administered. The model based on a level of precision of SEM < 0.55 was selected in this work but this choice can be criticized in terms of measurement precision, especially for decisions on individual test scores [78, 79]. This problem in terms of measurement precision is generally found in unidimensional and multidimensional item banks, and test developers need to consider and improve this issue in future works.

The missing data mechanism in this study was assumed to be missing at random (MAR). Multiple imputation was the only method available in the R package *mirt* [46] to impute missing data for estimating latent traits. Multiple imputation is a widely recommended method in the case of MAR data [80]. Moreover, to the best of our knowledge, there is no statistical test to accept or reject the hypothesis of MAR missing data. According to a previous study, MAR appears to be a reasonable approach for considering the missing data mechanism in the analysis of QoL data [81].

We did not use any content balancing algorithms in the item selection. However, the content balancing strategy may help the SQoL-MCAT to better balance the selection of items and achieve improved measurement accuracy. This issue is thus an important perspective to consider in future studies on the SQoL-MCAT. However, all patients had at least one item from each dimension administered.

The choice of the cut-off point in assessing the root mean square error (i.e., RMSE = 0.3) was based on a previous study comparing a CAT score with a full item bank score of depressive symptoms [64]. This cut-off may be not relevant in our study and in particular may lack sensitivity. However, the underlying assumption when presenting the RMSE is that the errors are unbiased and follow a normal distribution. In our study, we used the RMSE to assess the precision of our CAT score regarding the full item bank score. Both these scores are standardized and have the same metric. In addition, the errors followed a normal distribution. This issue should be studied in the future.

Validity is considered present when the measurement predicts an external criterion based on a gold standard. In the case of QoL, there is no gold standard, and as a consequence we cannot explore convergent validity. Future studies should, however, include other QoL measurements related to the concept measured by the SQoL to specifically evaluate convergent validity.

Responsiveness or sensitivity to change was not tested in our study. This parameter, defined as the ability to detect a meaningful change, is a core psychometric property of measurement instruments [73] and is of major interest in the follow-up of patients with schizophrenia in clinical practice and in therapeutic trials. The sensitivity to change should thus be confirmed in the SQoL-MCAT in future longitudinal studies.

Lastly, we considered that all the items of the SQoL were effect or reflective indicators and were thus allowed to be used in a latent trait model. However, this assertion is theoretical and future works should confirm empirically whether all the items of the SQoL are reflective [38].

Conclusion

The SQoL-MCAT is the first computerized adaptive QoL questionnaire specifically for patients with schizophrenia. As it is tailored to patient characteristics and is significantly shorter than the paper-based version, the SQoL-MCAT may improve the feasibility of assessing QoL in clinical practice, making it less burdensome to patients and allowing health professionals to obtain QoL data in real time.

Appendix

See Table 5.

Table 5 Differential item functioning according to gender, presence of paranoid symptoms, and insight

Item No.	Gender (male vs. female)		Presence of paranoid symptoms (paranoid schizophrenia vs. other)		Level of insight (aware vs. unaware)	
	p value	ΔR^2	*p* value	ΔR^2	*p* value	ΔR^2
1	0.9414	0.0001	0.4080	0.0015	0.9911	0.0004
2	0.6618	0.0005	0.3581	0.0017	0.2167	0.0091
3	0.3713	0.0013	0.6295	0.0008	**0.0000**	**0.0493**
4	0.3535	0.0013	0.2376	0.0024	0.6983	0.0034
5	0.0608	0.0035	0.1792	0.0028	0.1535	0.0103
6	0.0948	0.0030	0.1013	0.0037	0.0772	0.0128
7	0.3629	0.0013	0.2639	0.0021	0.6074	0.0041
8	0.2801	0.0016	0.9532	0.0001	0.2019	0.0090
9	0.9018	0.0001	0.1350	0.0033	0.7721	0.0028
10	0.7074	0.0005	0.4887	0.0013	0.3523	0.0069
11	0.0628	0.0035	0.3276	0.0018	0.3929	0.0063
12	0.6625	0.0006	0.5628	0.0010	0.2160	0.0092
13	0.7758	0.0003	0.4080	0.0014	0.6336	0.0039
14	0,7758	0,0003	0,9018	0,0001	0.7721	0.0028
15	0.2221	0.0020	0.5678	0.0009	0.7721	0.0028
16	0.3630	0.0013	0.3316	0.0018	0.3523	0.0069
17	0.0770	0.0033	0.5500	0.0010	0.6336	0.0039
18	**0.0113**	**0.0060**	0.7847	0.0004	0.9014	0.0017
19	0.7416	0.0004	0.6912	0.0006	0.8456	0.0024
20	0.6890	0.0005	0.7430	0.0005	0.3882	0.0073
21	0.6095	0.0007	0.4174	0.0015	0.5814	0.0049
22	0.8815	0.0002	0.4189	0.0015	0.9014	0.0017
23	0.4966	0.0010	0.3817	0.0019	0.4435	0.0069
24	**0.0000**	**0.0177**	0.6172	0.0008	0.5598	0.0053
25	0.9598	0.0001	0.3363	0.0018	0.7642	0.0029
26	0.5026	0.0009	0.2421	0.0023	**0.0011**	**0.0293**
27	0.9846	0.0000	**0.0001**	**0.0161**	0.5836	0.0048
28	0.1435	0.0027	0.4202	0.0015	0.4595	0.0060
29	0.1050	0.0030	0.1345	0.0033	0.1131	0.0123
30	0,8815	0,00002	0,7074	0,0005	0.3523	0.0069
31	0.1924	0.0021	0.2094	0.0025	0.4253	0.0059
32	0.0697	0.0034	0.6689	0.0006	0.2433	0.0083
33	**0.0019**	**0.0088**	0.4281	0.0015	0.4634	0.0058
34	0.4095	0.0012	0.1167	0.0038	0.3189	0.0076
35	0.1980	0.0021	0.9415	0.0001	0.3440	0.0071
36	0.1595	0.0024	0.2841	0.0020	0.5863	0.0044
37	0.4338	0.0011	0.5398	0.0011	0.1098	0.0121
38	0.1159	0.0029	0.1391	0.0035	0.1556	0.0109
39	0.5420	0.0008	0.0979	0.0040	0.7746	0.0028
40	0.5948	0.0007	0.3047	0.0020	0.8712	0.0019
41	0.2658	0.0018	0.7437	0.0005	0.0784	0.0154

Pierre Michel, Karine Baumstarck, Christophe Lancon, Badih Ghattas, Anderson Loundou, Pascal Auquier, Laurent Boyer (2017). Modernizing quality of life assessment: development of a multidimensional computerized adaptive questionnaire for patients with schizophrenia. Quality of Life Research

Bold values: DIF *p* value <0.05

Δ-R^2 DIF magnitude: negligible (Δ-$R^2 < 0.13$), moderate ($0.13 \leq \Delta$-$R^2 \leq 0.26$), or large (Δ-$R^2 \geq 0.26$)

References

1. Awad, A. G., Voruganti, L. N., & Heslegrave, R. J. (1997). Measuring quality of life in patients with schizophrenia. *PharmacoEconomics, 11*(1), 32–47.
2. Bobes, J., García-Portilla, P., Sáiz, P. A., Bascarán, T., & Bousoño, M. (oct 2005). Quality of life measures in schizophrenia. *European Psychiatry, 20*(Suppl 3), S313–S317.
3. Boyer, L., Baumstarck, K., Boucekine, M., Blanc, J., Lançon, C., & Auquier, P. (2013). Measuring quality of life in patients with schizophrenia: An overview. *Expert Review of Pharmacoeconomics & Outcomes Research, 13*(3), 343–349.
4. Greenhalgh, J., Long, A. F., & Flynn, R. (1982). The use of patient reported outcome measures in routine clinical practice: Lack of impact or lack of theory? *Social Science & Medicine, 60*(4), 833–843.
5. Morris, J., Perez, D., & McNoe, B. (1998). The use of quality of life data in clinical practice. *Quality of Life Research, 7*(1), 85–91.
6. Awad, A. G., & Voruganti, L. N. P. (2012). Measuring quality of life in patients with schizophrenia: An update. *PharmacoEconomics, 30*(3), 183–195.
7. Halyard, M. Y., Frost, M. H., Dueck, A., & Sloan, J. A. (2006). Is the use of QOL data really any different than other medical testing? *Current Problems in Cancer, 30*(6), 261–271.
8. Halyard, M. Y., Frost, M. H., & Dueck, A. (2006). Integrating QOL assessments for clinical and research purposes. *Current Problems in Cancer, 30*(6), 319–330.
9. Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., & Reeve, B., et al. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group during its first two years. *Medical Care, 45*(5 Suppl 1), S3–S11.
10. Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., & Yount, S., et al. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology, 63*(11), 1179–1194.
11. Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists* (1 ed). Mahwah, NJ: Psychology Press.
12. Fayers, P., & Machin, D. (2007). *Quality of life: The assessment, analysis and interpretation of patient-reported outcomes* (2e éd). New York: Wiley.
13. Weiss. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development, 37*, 70–84.
14. Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., & Teresi, J. A., et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care, 45*(5 Suppl 1), S22–S31.
15. Hill, C. D., Edwards, M. C., Thissen, D., Langer, M. M., Wirth, R. J., & Burwinkle, T. M., et al. (2007). Practical issues in the application of item response theory: A demonstration using items from the pediatric quality of life inventory (PedsQL) 4.0 generic core scales. *Medical Care, 45*(5 Suppl 1), S39–S47.
16. Gardner, W., Kelleher, K. J., & Pajer, K. A. (2002). Multidimensional adaptive testing for mental health problems in primary care. *Medical Care, 40*(9), 812–823.
17. Petersen, M. A., Groenvold, M., Aaronson, N., Fayers, P., Sprangers, M., Bjorner, J. B. (2006). Multidimensional computerized adaptive testing of the EORTC QLQ-C30: Basic developments and evaluations. *Quality Life Research, 15*(3), 315–329.
18. Haley, S. M., Ni, P., Dumas, H. M., Fragala-Pinkham, M. A., Hambleton, R. K., & Montpetit, K., et al. (2009) Measuring global physical health in children with cerebral palsy: Illustration of a multidimensional bi-factor model and computerized adaptive testing. *Quality Life Research, 18*(3), 359–370.
19. Auquier, P., Simeoni, M. C., Sapin, C., Reine, G., Aghababian, V., & Cramer, J., et al. (2003). Development and validation of a patient-based health-related quality of life questionnaire in schizophrenia: The S-QoL. *Schizophrenia Research*, 63(1–2), 137–149.
20. Baumstarck, K., Boyer, L., Boucekine, M., Aghababian, V., Parola, N., & Lançon, C., et al. Self-reported quality of life measure is reliable and valid in adult patients suffering from schizophrenia with executive impairment. *Schizophrenia Research, 147*(1), 58–67.
21. Baumstarck, K., Boucekine, M., Boyer, L., Aghababian, V., Parola, N., & Reuter, F., et al. (2014). Quantification of relevance of quality of life assessment for patients with cognitive impairment: The suitability indices. *BMC Neurology, 14*(1), 78.
22. Auquier, P., Tinland, A., Fortanier, C., Loundou, A., Baumstarck, K., & Lancon, C., et al. (2013). Toward meeting the needs of homeless people with schizophrenia: The validity of quality of life measurement. *PloS ONE, 8*(10), e79677.
23. Boyer, L., Lançon, C., Baumstarck, K., Parola, N., Berbis, J., & Auquier, P. (2013). Evaluating the impact of a quality of life assessment with feedback to clinicians in patients with schizophrenia: Randomised controlled trial. *The British Journal of Psychiatry, 202*, 447–453.
24. Catherine, F.-A., Boyer, L., Jonathan, W., Jean-Philippe, R., Raphaelle, R., Elisabeth, S., et al. (2015). Neural substrate of quality of life in patients with schizophrenia: A magnetisation transfer imaging study. *Scientific Reports, 5*, 17650.
25. Boyer, L., Richieri, R., Faget, C., Padovani, R., Vaillant, F., Mundler, O., et al. (2012). Functional involvement of superior temporal sulcus in quality of life of patients with schizophrenia. *Psychiatry Research, 202*(2), 155–160.
26. McKenna, S. P. (1997) Measuring quality of life in schizophrenia. *European Psychiatry, 12*(Suppl 3), 267s–274s.
27. Calman, K. C. (1984). Quality of life in cancer patients–an hypothesis. *Journal of medical ethics, 10*(3), 124–127.
28. Caqueo-Urízar, A., Boyer, L., Boucekine, M., & Auquier, P. (2014). Spanish cross-cultural adaptation and psychometric properties of the Schizophrenia Quality of Life short-version questionnaire (SQoL18) in 3 middle-income countries: Bolivia, Chile and Peru. *Schizophrenia Research, 159*(1), 136–143.
29. Chou, C.-Y., Ma, M.-C., Yang, T.-T., Chen-Sea, M.-J. (2011). Psychometric validation of the S-QoL Chinese (Taiwan) version for patients with schizophrenia. *Quality Life Research, 20*(5), 763–767.
30. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed., text rev.)*. Washington, DC: American Psychiatric Association.
31. Loi n° 78–17 du 6 janvier. (1978). relative à l'informatique, aux fichiers et aux libertés.
32. WMA Declaration of Helsinki - Ethical principles for medical research involving human subjects. (2013). Retrieved March 18, 2016, from http://www.wma.net/fr/30publications/10policies/b3/.
33. Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin, 13*(2), 261–276.
34. Lançon, C., Reine, G., Llorca, P. M., & Auquier, P. (1999). Validity and reliability of the French-language version of the Positive and Negative Syndrome Scale (PANSS). *Acta Psychiatrica Scandinavica, 100*(3), 237–243.

35. Addington, D, Addington, J, & Schissel, B. (1990) A depression rating scale for schizophrenics. *Schizophrenia Research, 3*(4), 247–251.
36. Michel, P., Baumstarck, K., Ghattas, B., Pelletier, J., Loundou, A., & Boucekine, M., et al. A multidimensional computerized adaptive short-form quality of life questionnaire developed and validated for multiple sclerosis. The MusiQoL-MCAT. *Medicine (Baltimore)*. 2016.
37. Fayers, P. M., & Hand, D. J.. (1997). Factor analysis, causal indicators and quality of life. *Quality Life Research, 6*(2), 139–150.
38. Costa, D. S. J. (2015). Reflective, causal, and composite indicators of quality of life: A conceptual or an empirical distinction? *Quality Life Research, 24*(9), 2057–2065.
39. Muthén, L., & Muthén, B. (2012). *MPlus user's guide* (7th ed). Los Angeles: Muthén & Muthén.
40. Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*(5), 893–898.
41. Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores*. Iowa City: Psychometric Society.
42. Samejima, F. (1974). Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika, 39*(1), 111–121.
43. Adams, R. J., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1–23.
44. Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care, 38*(9 Suppl), II28-II42.
45. Cai L. (2010). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics, 35*(3), 307–335.
46. Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48, 1–29.
47. Kang, T., & Chen, T. T. (2008). Performance of the generalized S-X2 item fit index for polytomous IRT models. *Journal of Educational Measurement, 45*(4), 391–406.
48. Raju, N. S., Price, L. R., Oshima, T. C., & Nering, M. L. (2007). Standardized conditional SEM: A case for conditional reliability. *Applied Psychological Measurement, 31*(3), 169–80.
49. Wang, T, & Others, A. (1996). Conditional standard errors, reliability and decision consistency of performance levels using polytomous IRT. Retrieved February 18, 2015, from http://eric.ed.gov/?id=ED401323.
50. Wright, B. D., & Linacre, J. M. (1994) Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.
51. Molenaar, I. W. (1991). A weighted Loevinger H-coefficient extending Mokken scaling to multicategory items. *Kwant Methoden, 12*(37), 97–117.
52. van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software, 39*(8), 1–30.
53. Stochl, J., Jones, P. B., & Croudace, T. J. (2012). Mokken scale analysis of mental health and well-being questionnaire item responses: A non-parametric IRT method in empirical research for applied health researchers. *BMC Medical Research Methodology, 12*, 74.
54. Siu, C. O., Harvey, P. D., Agid, O., Waye, M., Brambilla, C., & Choi, W.-K., et al. (2015). Insight and subjective measures of quality of life in chronic schizophrenia. *Schizophrenia Research: Cognition, 2*(3), 127–132.
55. Boyer, L., Cermolacce, M., Dassa, D., Fernandez, J., Boucekine, M., & Richieri, R., et al. (2012). Neurocognition, insight and medication nonadherence in schizophrenia: A structural equation modeling approach. *PloS ONE, 7*(10), e47655.
56. Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R Package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*(8), 1–30.
57. Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores. Ottawa, Directorate of Human Resources Research and Evaluation, Department of National Defense.
58. Anatchkova, M. D., Saris-Baglama, R. N., Kosinski, M., & Bjorner, J. B. (2009). Development and preliminary testing of a computerized adaptive assessment of chronic pain. *The Journal of Pain, 10*(9), 932–943.
59. Huisman, M., & Molenaar, I. W. (2001). Imputation of missing scale data with item response models. In: Boomsma, A., van Duijn, M. A. J. & Snijders, T. A. B. (Eds.) *Essays on item response theory*, pp. 221–244. New York: Springer. (Lecture Notes in Statistics). Retrieved September 20, 2013, from http://link.springer.com/chapter/10.1007/978-1-4613-0169-1_13.
60. Peyre, H., Leplège, A., & Coste, J. (2010). Missing data methods for dealing with missing items in quality of life questionnaires. A comparison by simulation of personal mean score, full information maximum likelihood, multiple imputation, and hot deck techniques applied to the SF-36 in the French 2003 decennial health survey. *Quality of Life Research, 20*(2), 287–300.
61. Mulder, J, & van der Linden, W. J.. (2009). Multidimensional adaptive testing with Kullback–Leibler information item selection. In: van der Linden, W. J. & Glas, C. A. W. (Eds.) *Elements of adaptive testing*, pp. 77–101. New York: Springer. (Statistics for Social and Behavioral Sciences). Retrieved January 28, 2015, from http://link.springer.com/chapter/10.1007/978-0-387-85461-8_4.
62. Harvill, L. M. (1991). Standard error of measurement. *Educational Measurement: Issues and Practice, 10*(2), 33–41.
63. Babcock, B., & Weiss, D. J. (2013). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement? *Journal of Computerized Adaptive Testing, 1*, 1–18.
64. Choi, S. W., Reise, S. P., Pilkonis, P. A., Hays, R. D., & Cella, D. (2010). Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. *Quality of Life Research, 19*(1), 125–136.
65. R Core Team. (2016). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.
66. Alessandrini, M., Lançon, C., Fond, G., Faget-Agius, C., Richieri, R., & Faugere, M., et al. (2016). A structural equation modelling approach to explore the determinants of quality of life in schizophrenia. *Schizophrenia Research, 171*(1–3), 27–34.
67. Ruggeri, M., Nosè, M., Bonetto, C., Cristofalo, D., Lasalvia, A., Salvi, G., et al. (2005). Changes and predictors of change in objective and subjective quality of life: Multiwave follow-up study in community psychiatric practice. *The British Journal of Psychiatry, 187*, 121–130.
68. Reine, G., Simeoni, M.-C., Auquier, P., Loundou, A., Aghababian, V., & Lancon, C. (2005). Assessing health-related quality of life in patients suffering from schizophrenia: A comparison of instruments. *European Psychiatry, 20*(7), 510–519.
69. Kemmler, G., Holzner, B., Neudorfer, C., Meise, U., & Hinterhuber, H. (1997). General life satisfaction and domain-specific quality of life in chronic schizophrenic patients. *Quality of Life Research, 6*(3), 265–273.
70. Folsom, D. P., Depp, C., Palmer, B. W., Mausbach, B. T., Golshan, S., & Fellows, I., et al. (2009). Physical and mental

health-related quality of life among older people with schizophrenia. *Schizophrenia Research, 108*(1–3), 207–213.
71. Smits, N. (2016). On the effect of adding clinical samples to validation studies of patient-reported outcome item banks: A simulation study. *Quality of Life Research, 25*(7), 1635–1644.
72. Yao, L. (2014). Multidimensional CAT item selection methods for domain scores and composite scores with item exposure control and content constraints. *Journal of Educational Measurement, 51*(1), 18–38.
73. Petersen, M. A., Giesinger, J. M., Holzner, B., Arraras, J. I., Conroy, T., & Gamper, E.-M., et al. (2013). Psychometric evaluation of the EORTC computerized adaptive test (CAT) fatigue item pool. *Quality of Life Research, 22*(9), 2443–2454.
74. Zhang, J. (2012). Calibration of response data using MIRT models with simple and mixed structures. *Applied Psychological Measurement, 36*(5), 375–398.
75. Baumstarck, K., Boyer, L., Boucekine, M., Aghababian, V., Parola, N., & Lançon, C., et al. (2013). Self-reported quality of life measure is reliable and valid in adult patients suffering from schizophrenia with executive impairment. *Schizophrenia Research, 147*(1), 58–67.
76. Girard, V., Tinland, A., Boucekine, M., Loubière, S., Lancon, C., & Boyer, L., et al. (2016). Validity of a common quality of life measurement in homeless individuals with bipolar disorder and schizophrenia. *Journal of Affective Disorders, 204*, 131–137.
77. Woods, C. M. (2014). Estimating the latent density in unidimensional IRT to permit non-normality. In *Handbook of item response theory modeling: Applications to typical performance assessment*. New York: Routledge.
78. Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement, 21*(4), 347–360.
79. Nunnally, J. C, & Bernstein, I. H. (1994) *Psychometric theory*, p. 736 (3rd ed). New York: McGraw-Hill.
80. Carpenter, J. R., & Kenward, M. G. (2008). *Missing data in randomised controlled trials—a practical guide*. Birmingham: The National Institute for Health Research. RM03/JH17/MK.
81. Fielding, S., Fayers, P. M., & Ramsay, C. R. (2009). Investigating the missing data mechanism in quality of life outcomes: A comparison of approaches. *Health and Quality of Life Outcomes, 7*, 57.
82. Nortvedt, M. W., Riise, T., Myhr, K. M., & Nyland, H. I. (2000). Quality of life as a predictor for change in disability in MS. *Neurology, 55*(1), 51–54.