Research Report

European Addiction Research

Eur Addict Res 2009;15:63–70
DOI: 10.1159/000189784

Published online: January 10, 2009

Development of a Specific Health-Related Quality of Life Test in Drug Abusers Using the Rasch Rating Scale Model

O.M. Lozano Rojas[a] A.J. Rojas Tejada[b] C. Pérez Meléndez[c]

[a]Psychology Department, University of Huelva, Huelva, [b]Social and Behavioral Science Department, University of Almeria, Almeria, [c]Social and Behavioral Science Department, University of Granada, Granada, Spain

Key Words

Health-related quality of life · Drug abusers · Rasch models · Rating scale model

Abstract

A test has recently been designed for the measurement of health-related quality of life in drug abusers (Health-Related Quality of Life for Drug Abusers Test – HRQOLDA Test) based on the bi-axial concept of addiction. The aim of this work is to find out the metric properties of the test by applying the polytomous Rasch Rating Scale Model (RSM). The HRQOLDA Test was given to a sample of 358 drug abusers who began treatment in a Therapeutic Community. The Rating Scale Model (RSM) and the WINSTEPS programme were used for the analysis of its metric properties. The results shown here are an overall fit of data to the model. The items are adequately distributed over the HRQOL continuum. The response categories appear to be in order except for the category 'a little'. The test region which measures most accurately, and is most informative, is in the middle of the continuum. The test properties studied confirm the adequacy of the RSM for accurate measurement of HRQOL in drug abusers, providing a test for measuring this specific construct in this population.

Introduction

The health-related quality of life (HRQOL) measure is widely used in the field of health. This construct measures individual responses to physical, mental and social effects of illness on daily life [1]. In the area of drug addiction, HRQOL has been used for a relatively short time. Reno and Aiken [2] recommend it for evaluating harm reduction programs, where drug consumers are considered chronic patients and the purpose of therapy goes beyond abstinence, and improvement of HRQOL is a basic goal of treatment.

HRQOL in the area of drug addiction has been measured with both generic and specific instruments. Generic instruments do not assess disease-specific symptoms. One drawback is that they may not be sensitive to significant clinical changes when specific diseases are evaluated. This limitation has been pointed out by some authors who have used them in drug abuse [3, 4]. They are therefore useful when the purpose of the study is to compare the HRQOL of drug abusers with other groups of patients or with the general reference population. They are also used when there is no specific instrument for a specific illness [5]. The generic tests most commonly used for HRQOL measurement in the drug addict population are the Nottingham Health Profile [3, 6], the SF-36 [7–9], WHOQOL-BREF [10, 11] and EQ-5D [12–14].

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1022–6877/09/0152–0063$26.00/0

Accessible online at:
www.karger.com/ear

Antonio J. Rojas Tejada
Social and Behavioral Science Department
University of Almeria
ES–04120 Almeria (Spain)
Tel. +34 9500 15538, Fax +34 9500 15420, E-Mail arojas@ual.es

The distinctive characteristic of instruments specific to HRQOL is that their content is adapted to the population subject of the study, and, therefore, this type of instrument is attributed greater sensitivity in detecting changes in the HRQOL. This makes them a likely instrument for studying efficacy and effectiveness, although not for comparing with other samples of subjects. In the area of drug addiction, the specialized literature reviewed shows that at present there is one specific instrument: Injection Drug User Quality of Life – IDUQOL [15–17].

The Health-Related Quality of Life for Drug Abusers Test (HRQOLDA Test) [18] is a quality-of-life instrument specific to drug abuse and is based on the bi-axial addiction concept of Edwards and Gross [19]. This instrument is based on an operational definition specific to drug abusers. It has been developed as a standardized measure of the effects of drug abuse on physical perception, psychological health and social functioning.

As a rule, the Classical Test Theory (CTT) has been the psychometric theory employed to measure HRQOL. However, the metric characteristics of this theory present a double invariance problem: (1) patient measurements depend on the instrument used (e.g. a patient would have different HRQOL scores depending on the test used), and (2) the estimates of the items and test properties depend on the sample of individuals used for this purpose (e.g. the reliability of a test depends on the sample of people used to calculate it). Furthermore, in the CTT we assume, (hardly credible though it may be) that once the reliability of a test has been estimated for a certain population, this reliability (accuracy) remains constant for all ability levels (e.g. it remains identical when estimating the measurement in persons with a high, medium or low HRQOL), whereas accuracy is most often lower when measuring the extremes of this continuum (high and low ability).

Advances in psychometrics have led to the CTT being displaced in favor of Item Response Theory (IRT) models. With these models, measurements are invariable, regardless of the instruments used and of the individuals evaluated. Calibration is independent of the sample to which the test is given (it is invariant over the population), and the measurements of persons are also test-free (it does not matter what selection of items is used to estimate them). Moreover, it is senseless to refer to test reliability where IRT is concerned, as measurement accuracy is estimated specifically (standard error of measurement) for each ability level in the variable. Other advantages over CTT include the additivity of items, and interval level measurement [20–25].

The main purpose of administering HRQOL tests is to assess individual ability in order to make a clinical interpretation. On the Rasch model, the items are in order of difficulty, and one of its greatest benefits is the possibility of mapping the test. This map provides individual item responses and ability estimates on each scale with criterion measures. When Rasch models are applied, the test should be able to identify the level of functioning, strengths and weaknesses, and implications for intervention.

For the reasons above, a Rasch IRT model, the Rating Scale Model (RSM), was the psychometric model employed to assess the metric properties of the HRQOLDA Test.

The purpose of this work was to assess the metric properties of the HRQOLDA Test developed from the operative definition above, by applying the Rating Scale Model. The metric properties studied are those of polytomous IRT models, analysis of data fit to the model, calibration of items, response category functioning and measurement accuracy of the various HRQOL levels.

Materials and Methods

Persons

The HRQOLDA Test was given to 358 patients admitted to five therapeutic communities run by the Andalusian Foundation for Drug Addiction Care. The therapeutic communities are located in several different provinces in Andalusia (Spain): 23.1% of the participants were from Almonte in Huelva, 29.2% from Cartaya in Seville, 16.7% from Los Palacios in Seville, 7.5% from Mijas in Malaga, and 23.4% were from Tarifa in Cadiz. 92.5% were men and 7.5% were women. The average age of the men was 34.7 (SD 7.9), and of the women 39.7 years (SD 8.2).

Participants were informed of the purpose of the study. They were explained that their participation was voluntary and that the data would be used exclusively for statistical purposes. At the end of the explanation, they were asked to sign an informed consent form for their participation.

The test was given by previously trained psychologists from the therapeutic communities in individual sessions with the patients during their first week in the centre.

Most of the patients were polydrug abusers. The drugs they had used most often in the month before being admitted were a mixture of heroine and cocaine base (31.4%), alcohol (25.8%), cocaine base (19%), cocaine hydrochloride (10.4%), cannabis (3.4%) and benzodiazepines (1.1%). 8.7% did not take drugs.

Instruments

The HRQOLDA Test was developed following the suggestions of the APA, AERA and NCME test standards [26]. The operative definition of the drug abuser HRQOL was developed in two stages. First, based on a review of the literature and expert judgment, the essential contents that should be included on a test measuring HRQOL were established. Once the basic contents had been defined, they were adapted to drug abusers for an operative definition of HRQOL specific to this population. Literature specialized in drug abuse was reviewed for this propose, and using the biaxial dependence concept as a reference, the most frequent harmful physical, psychological and behavioral symptoms in drug users were included. This operative HRQOL definition specific to drug abusers was evaluated by experts in drug abuse. The final operative definition is made up of a physical category that includes the person's physical functional status, symptoms and dependence. A second, psychosocial category includes psychological dependence, symptoms, deterioration of cognitive functioning, general perception of health, social functioning, and expectations [18].

Five items were chosen for each point in the operative definition content and two psychometry experts proceeded to analyse them for item-objective congruence. Afterwards, a qualitative analysis of the items selected was made by giving them to a sample of drug abuse patients. The purpose of this stage was to find out whether the items were understandable to the drug abuser population, and to evaluate the comprehension of the test instructions and the easiest response format. After the qualitative analysis, the item for each point in the content that had worked the best was selected, and a preliminary version of the test was created. This preliminary version was given to a sample of 50 patients, all of the items showing a suitable response distribution in the various choices, and adequate discrimination indices. The final version of the HRQOLDA Test is made up of 20 five-choice scaled-response items.

The choices are scored as follows: (a) on positive items: 'Not at all' (5 points), 'A little' (4 points), 'Sometimes' (3 points), 'Quite a bit' (2 points) and 'A lot' (1 point), and (b) on negative items (items 15 and 19): 'Not at all' (1 point), 'A little' (2 points), 'Sometimes' (3 points), 'Quite a bit' (4 points) and 'A lot' (5 points). A higher score on the test represents better quality of life.

Finally, validation studies carried out with the HRQOLDA Test have shown sensitivity to changes in HRQOL. This test has shown sensitivity to change in pre-post treatment studies and discriminates between patients in different stages of treatment [18]. Multiple regression analysis reveals that HRQOL Test scores can be predicted from drug dependence, physical health status and psychological adjustment [27].

Rasch Model

The metric properties of the HRQOLDA Test were found using the RSM [28–32], an extension of the Rasch model [33] suitable for rating scale data. This model assumes a functional relationship between the variable measured by the items on a test and the person's responses to them. The function that determines the probability person n has of answering in a certain response category x on item i (β_{nix}) is given by the following logistic function [28–32]:

$$\pi_{nix} = \frac{\exp\sum_{j=1}^{x}\left[\beta_n - \left(\delta_i + \tau_j\right)\right]}{\sum_{x=1}^{m+1}\exp\sum_{j=1}^{x}\left[\beta_n - \left(\delta_i + \tau_j\right)\right]} \qquad x = 1, 2, \ldots, m+1$$

where β_n measures patient n's ability to achieve a high score (i.e. high HRQOL); δ_i is the item's calibrated 'difficulty' (to what extent an item represents a high HRQOL), and τ_j, called the threshold parameter, is the location of threshold m with respect to item location on the scale. The important quantity modeled is the difference between patient ability and item 'difficulty' plus the threshold parameter 'difficulty' [$\beta_n - (\delta_i + \tau_j)$]. Therefore, persons with different abilities in the variable measured will have different probabilities of scoring on a given item.

The decisive property of Rasch models is that a person's abilities, item and threshold parameter are estimated independently by means of conditional maximum likelihood. This means that the probability of response to an item depends only on the ability of the person in the latent trait evaluated.

To interpret ability, both in measuring persons and item calibration, data are logarithmically transformed so that both remain on an interval scale called the 'logit scale' (log-odds units), with 0 mean and a SD of 1 [32].

The WINSTEPS programme version 3.64.2 [34] was used to apply the RSM.

Analysis of Fit

Residual analysis, which tests the degree to which the test response data are as expected from the model, was used as the index of fit by analyzing the differences between observed and expected values for a person with a certain ability on each item in the test [32, 35]. For this analysis, the continuum is divided into K intervals, and the percentage of correct responses Pjk and the percentage of responses expected according to the model are evaluated for each interval E(Pjk).

Two fit statistics were analyzed, the mean-square (MNSQ) and the standardized mean-square (ZSTD) [36]. The latter takes into account sample error and follows a normal distribution with a 0 mean and a SD of 1. There are, in turn, two MNSQ and ZSTD statistics called infit and outfit. The first is an information-weighted fit statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's level of ability. The outfit is an outlier-sensitive fit statistic, more sensitive to unexpected behavior on items far from the person's level of ability. The region considered an acceptable fit in the MNSQ varies between 0.6 and 1.4 where 1 is ideal [36, 37]. In the ZSTD, 0 indicates concordance between observed and model data. Linacre [38] suggests that the interval between –2 and +2 indicates an acceptable fit.

Interpretation of the 'HRQOL' and item calibration was done by transforming data to the 'logit' scale, which describes a probabilistic relationship between item difficulty (δi) and a person's ability (θn). For any person and any item, the difference (θn – δi) is a magnitude with a uniform meaning across the scale [32].

Results

Fit between Data and Model

Fit can be studied from three perspectives: (1) the overall or total fit (which enables the data fit to the overall model to be assessed); (2) person fit (which helps identify a suspicious pattern of responses), and (3) item fit (which helps identify poor items). For the purposes of this study, we focus on sections 1 and 3. Section 2, although extremely important, is not examined in this study, since

Table 1. Summary of persons, items and item statistics

	n	Measure	SE	Infit		Outfit		r_{pb}
				MNSQ	ZSTD	MNSQ	ZSTD	
Summary of persons								
Mean		0.27	0.21	1.01	–0.1	1.03	–0.1	
SD		0.62	0.05	0.48	1.0	0.56	1.0	
Maximum		3.30	0.93	2.90	2.6	4.18	3.1	
Minimum		–1.41	0.19	0.25	–2.4	0.26	–2.3	
Summary of items								
Mean		0.0	0.05	1.02	0.0	1.03	0.0	
SD		0.57	0.00	0.23	1.0	0.31	1.0	
Maximum		1.26	0.06	1.47	1.6	2.00	2.3	
Minimum		–1.00	0.05	0.65	–1.8	0.65	–1.7	
Item statistics								
01 Physical pain	354	–0.11	0.05	0.76	–1.2	0.80	–0.9	0.53
02 Insomnia	358	0.22	0.05	1.01	0.0	1.01	0.0	0.50
03 Nauseas	358	–0.64	0.05	1.1	0.4	0.99	0.0	0.46
04 Energy loss	358	0.29	0.05	0.65	–1.8	0.65	–1.6	0.62
05 Nervousness	358	0.65	0.05	0.68	–1.5	0.67	–1.5	0.64
06 Depression	357	0.55	0.05	0.75	–1.1	0.81	–0.8	0.59
07 Aggressiveness	358	–0.11	0.05	0.94	–0.3	0.93	–0.3	0.48
08 Memory problems	358	0.13	0.05	0.94	–0.3	0.91	–0.4	0.55
09 Hallucinations	358	–0.80	0.06	1.23	0.8	1.11	0.3	0.45
10 Concentration problems	358	0.01	0.05	0.67	–1.7	0.68	–1.5	0.60
11 Orientation problems	358	–1.00	0.06	1.06	0.2	0.89	–0.3	0.47
12 Physical dependence	358	0.20	0.05	1.27	1.1	1.26	1.0	0.48
13 Psychological dependence	357	0.22	0.05	1.30	1.2	1.31	1.2	0.44
14 Functional activities	357	–0.35	0.05	1.03	0.1	0.96	–0.2	0.55
15 Emotional support	357	–0.69	0.05	1.42	1.6	2.00	2.3	0.01
16 Subjective psychological perception	356	0.00	0.05	0.84	–0.8	0.83	–0.7	0.54
17 Family role	357	0.35	0.05	1.31	1.3	1.38	1.4	0.27
18 Social role	357	1.26	0.06	1.21	0.7	1.15	0.4	0.38
19 Self-referential expectations	358	–0.83	0.06	1.09	0.3	1.28	0.7	0.15
20 Subjective physical perception	358	0.64	0.05	1.00	0.0	0.98	–0.1	0.50

Threshold parameters: $\tau_1 = -0.58$; $\tau_2 = -0.14$; $\tau_3 = 0.76$; $\tau_4 = -0.04$.

our focus is on the analysis of the metric properties of the HRQLDA Test, not the analysis of response patterns of the individuals.

In table 1, the first two rows show total fit, OUTFIT (MNSQ and ZSTD) and INFIT (MNSQ and ZSTD) for persons and items. The results found for persons (MNSQ INFIT mean = 1.01; ZSTD INFIT mean = –0.1 and SD = 1.0; MNSQ OUTFIT mean = 1.03; ZSTD OUTFIT mean = –0.1 and SD = 1.0), and items (MNSQ INFIT mean = 1.02; ZSTD INTFIT mean = 0 and SD = 1.0; MNSQ OUTFIT mean = 1.03; ZSTD OUTFIT mean = 0 and SD = 1.0), support the interpretation of adequate total fit of persons and items (table 1). 88.9% of persons fit to the model.

Fit of items in the table shows that 19 of 20 items fit to the model prediction. Only the parameters of Item 15 have misfitting OUTFITS, in which a low point-biserial correlation is observed.

Calibration of Items and Measurement of Persons

The persons and items map (fig. 1) shows their position by HRQOL level, measured on a logit scale. In the persons column (left), each '#' represents 3 participants and each '.' 1 or 2 subjects. The position of the items on the scale appears on the right, with their value in parentheses (table 1). The items/persons highest on the scale are at the top of the graph, while those items/persons lowest on the scale are at the bottom. The letters 'M', 'S' and

'Q' for items and persons stand for the mean, SD and 2 SDs, respectively.

The items are between 1.26 logits (Item 18: social role) and –1.0 logits (Item 11: orientation) on the scale. The distances observed between items show adequate distribution over the HRQOL continuum. Only the distance of 0.59 logits between Items 5 (anxiety) and 18 (social role) is a little wider than recommended [39]. There is no overrepresentation of items in the construct; although some items are closer together, the content they represent is different.

Characteristic Curves of Response Categories

In the RSM, the ability threshold parameters for the item thresholds (τ_m) are the same across all items (although not between response categories), and the difference between items resides in the position they occupy on the continuum. Their use is indicated when all of the test items are defined by the same response categories [40].

In the HRQOLDA Test, all items are defined by the same categories (1, 2, 3, 4 and 5), and have four threshold parameters. Table 1 shows the parameter values. It may be observed that the category threshold parameters appear to be in order, except for the category 'A little' which is out of order with respect to the adjacent categories. This shows that in this category the empirical order is inconsistent with the theoretical order proposed [41].

Characteristic curves are defined using the response category threshold parameters that make it possible to visualize the regions of the continuum where the response is most likely.

As shown in figure 2, the categories 'A lot', 'Quite a bit', 'Sometimes' and 'Not at all' delimit four regions on the continuum where the responses of the subjects are more likely. The category 'A little' is not the most likely answer in any region on the continuum because its threshold parameter is higher than for the category 'Sometimes', which suggests the advisability of using only four item response categories.

Measurement Accuracy

Figure 3 shows the HRQOLDA Test information function, which makes it possible to examine test accuracy at different HRQOL levels. It may be observed that the test is most accurate in the middle between –0.5 and 0.5 logits, where 61.3% of the subjects scored; 86.4% of the subjects are between –1 and 1 logits, and 96.4% scored between –1.5 and 1.5 logits.

The index of separation between persons was 2.64, corresponding to a reliability of 0.85, and the index of

```
PERSONS              MAP OF ITEMS
        <more HRQoL>|<more difficult>
  2              .  +
                 .  |
                    |
                 #  |
                 .  |
                ## Q|
                ##  |
                 #  |  i_18 (1,26)
               .##  |
                .#  |Q
  1           .###  +
              .### S|
               ###  |
              .###  |  i_5 (0,65) i_20 (0,64)
          ########  |S i_6 (0,55)
            .#####  |
        .#########  |
   .############ M|  i_17 (0,35) i_4 (0,29)
           .######  |  i_12 (0,2) i_13 (0,22) i_2 (0,22)
          .#######  |  i_8 (0,13)
  0      .########  +M i_10 (0,1) i_16 (0)
        .#########  |  i_1 (-0,11)   i_7 (-0,11)
             .####  |
            .#####  |
              .### S|  i_14 (-0,35)
              .###  |
                .#  |S i_3 (-0,64)
                .#  |  i_15 (-0,69)
               .##  |  i_19 (-0,83) i_9 (-0,8)
                .#  |
 -1              . Q+  i_11 (-1.0)
                 .  |Q
                 .  |
                 .  |
                 .  |
                    |
                    |
                    |
                    |
                    |
 -2                 +
        <less HRQoL>|<less difficult>
```

Fig. 1. Distribution map for persons and items.

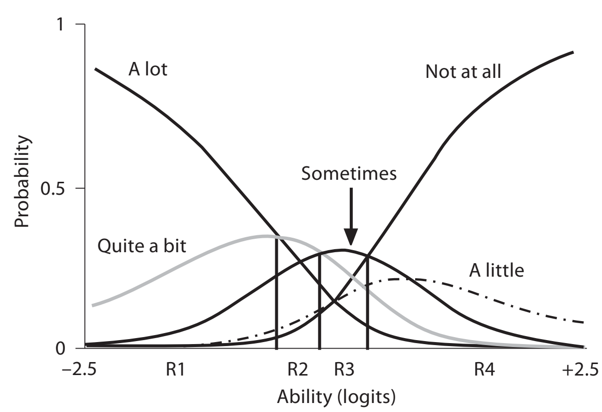

Fig. 2. Characteristic curves of item response categories.

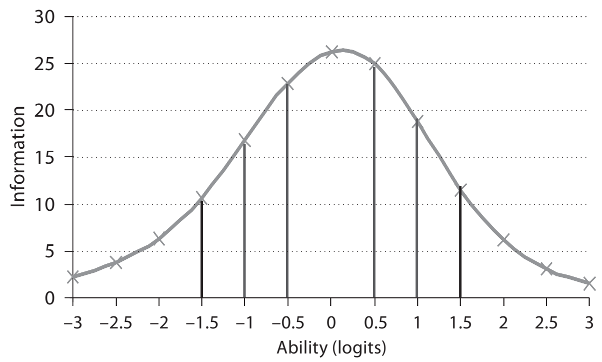

Fig. 3. Test information function.

separation between items was 11.31, corresponding to a CTT reliability of 0.99. The mean SE of measurement of the items is 0.05 logits.

Discussion

This paper presents the HRQOLDA Test for measuring HRQOL. This test contributes two important innovations, the first of which is highly transcendent for clinical practice and the second entails great metric advantages.

The first contribution is that the HRQOLDA Test satisfies the characteristics required for an HRQOL test specific to the drug user population. The use of the HRQOLDA Test, compared to generic tests that measure HRQOL, provides an advantage of enormous practical usefulness to the clinical professional. Generic tests (with nonspecific content for patients with any health problem) are versatile enough to be applied to samples of patients with very diverse problems. This may be an advantage in certain contexts, but represents a limitation when a more exhaustive evaluation of the HRQOL is required, such as in a population with a specific problem. In the sphere of drug abuse, several authors have pointed this out in their studies [3, 4] and indicate the need to have HRQOL tests designed for the drug abuse population. The HRQOLDA Test represents an innovation over generic HRQOL tests currently in use as the contents the items are based on come from an analysis of the physical, psychological and social symptoms most often deteriorating the daily life of drug users. Therefore, compared to generic HRQOL tests, this test's scores show contents more susceptible to change (improvement or deterioration) in drug abuse patients. It should be remembered that the use of HRQOL as a measure of treatment effectiveness requires highly sensitive tests.

The second contribution is that the HRQOLDA Test was developed using the RSM, a Rasch IRT model. The application of IRT models has a series of advantages: (1) the measure of persons and items together on the same scale (logits). Based on these models, the scale or ability can be found not only for persons, but for items as well, which is very useful for their analysis. (2) IRT polytomous models show how items and response categories work. There are also some disadvantages in IRT models compared to the CTT, which in spite of its strong limitations, is simple (in terms of easily calculated metric properties), flexible, and very well known. The IRT solves many CTT deficiencies, but its models are more complex (requiring more complicated calculations and depending on software for its application), more rigid (if the data do not fit to the model, measurement is not possible) and little known (at least in applied fields). These issues have hindered more extensive use of IRT.

In this research paper, we have shown that the RSM is a feasible and practicable way of approaching HRQOL measurement by studying the metric properties of the HRQOLDA Test. This should cause us to reflect on the possibility of using IRT-based models instead of classic approaches (based on CTT), above all, considering their advantages (measurement invariance, measurement of persons and items together, study of how response categories operate, etc.). In this study, test accuracy demonstrated that the parameters in the person and item reliability indices are suitable. It was also observed that over half of the persons interviewed (59.7%) had scores in the region where the test is most accurate (between –0.5 and 0.5 logits), and 92.7% of the sample scored in a region on the continuum where the test also proved to be highly accurate (between –1 and 1 logits). That is, professionals have clinical assurance that these subjects have been accurately measured. On the other hand, the test is less accurate for the remaining 7.3% of subjects. Therefore, for an adequate estimate of the HRQOL of these patients, clinicians will have to perform other complementary tests.

IRT models make it possible to minimize floor and ceiling effects of tests in the field of health [42]. The person and item map shows that the items on the HRQOLDA test are adequately distributed along the HRQOL continuum. 91.9% of the persons have an HRQOL between the highest and lowest on the scale. Calibration of the 20 items shows that 10 are over 0 logits on the scale, 8 items are less than 0 logits and two are in the centre of the continuum.

According to the results found in this study and in other validation studies already performed [18, 27], it may be said that in general, the test shows psychometric properties suitable for measuring the HRQOL in a drug user population. However, as commented above, the test has 2 weaknesses.

On the one hand, Item 15 did not adjust adequately to the model. The authors consider two hypotheses that might explain the lack of fit of Item 15, which measures emotional support: (a) The content of the item is not appropriate to the study sample, which is made up of persons who have usually gone into treatment because their family and friends convinced them. That is, almost all of the persons that make up the sample have emotional support. In this case, the item would not discriminate well enough and could lead to the lack of fit observed. (b) The item was not well written and shows a social desirability bias. This bias does not meet the one-dimensionality assumption of the Rasch models, and the answers would therefore vary due to the content the item measures and to the social desirability bias.

Although Item 15 does not fit, it is recommended that it be not eliminated because of the importance its content has to the operative definition of the HRQOL in the HRQOLDA Test.

On the other hand, study of the characteristic response category curves shows inadequate behavior of the choice 'a little'. According to the model, it is more likely that a person would respond to the category 'sometimes' than to the category 'a little'. Therefore, this response category might be eliminated and only four categories per item used. This would have to be approached in a new study.

In conclusion, this work has used a Rasch model to show that the metric properties of the HRQLDA Test are suitable for use. The results allow its use to be recommended mainly for clinical research, where the purpose is to detect changes in patients. The HRQLDA Test is predictably more sensitive than generic tests for evaluating the effect of treatments. On the other hand, generic instruments still seem to be better for epidemiological studies in which it is desired to compare the HRQOL of patients with each other or with the general population.

We think new validation studies should be done to compare the application of this test with the generic tests used up to now, such as the SF-36 and the NHP, and find empirical evidence on their similarities and differences.

Acknowledgments

Research was conducted at the University of Almeria in Spain, with funding from the 2003–2006 Andalusian Plan for Research, Technological Development and Innovation. The authors would like to thank Deborah Fuldauer for her assistance with the English language in the manuscript.

References

1 Bowling A: Measuring Illness: A Review of Disease Specific Quality of Life Measurement Scales. Philadelphia, Open University Press, 1994.

2 Reno RR, Aiken LS: Life activities and life quality of heroin addicts in and out of methadone treatment. Int J Addict 1993;28:211–232.

3 Torrens M, San L, Martínez A, Castillo C, Domingo-Salvany A, Alonso J: Use of the Nottingham Health Profile for measuring health status of patients in methadone maintenance treatment. Addiction 1997;92:707–716.

4 Vanagas G, Padaiga Z, Subata E: Drug addiction maintenance treatment and quality of life measurement. Medicina (Kaunas), 2004; 40:833–841.

5 Badía X: La Investigación de resultados en salud: de la evidencia a la práctica clínica. Barcelona, Edimac, 2000.

6 Puigdollers E, Domingo-Salvany A, Brugal MT, Torrens M, Alvaros J, Castillo C, Magri N, Martin S, Vazquez JM: Characteristics of heroin addicts entering methadone maintenance treatment: quality of life and gender. Subst Use Misuse 2004;39:1353–1368.

7 Daeppen JB, Krieg MA, Burnand B, Yersin B: MOS-SF-36 in evaluating health-related quality of life in alcohol-dependent patients. Am J Drug Alcohol Abuse, 1998;24:685–694.

8 O'Brien S, Mattick R, White J, the NEPOD Research Group: Maintenance pharmacotherapy for opioid dependence and SF-36 health status: a comparison with general population norms and other chronic disorders. Addict Dis Their Treatm 2006;5:155–164.

9 Villeneuve PJ, Challacombe L, Strike CJ, Myers T, Fischer B, Shore R, Hopkins S, Millson P: Change in health-related quality of life of opiate users in low-threshold methadone programs. J Subs Use 2006;11:137–149.

10 Barros da Silva A, Fleck M, Pechansky F, de Boni R, Sukop P: Psychometric properties of the World Health Organization Quality of Life instrument (WHOQoL-BREF) in alcoholic males: a pilot study. Qual Life Res 2005; 14:473–478.

11 Bizzarri UJ, Rucci P, Vallotta A, Girelli M, Scandolaru A, Zerbetto E, Sbrana A, Iagher C, Dellantonio E: Dual diagnosis and quality of life in patients in treatment for opioid dependence. Subst Use Misuse 2005;40:1765–1776.

12 Foster JH, Peters TJ, Marshall EJ: Quality of life measures and outcome in alcohol-dependent men and women. Alcohol 2000;22:45–52.

13 Franic DM, Jiang JZ: Potentially inappropriate drug use and health-related quality of life in the elderly. Pharmacotherap*y* 2006;26: 768–778.

14 Günther O, Roick C, Angermeyer MC, König HH: The EQ-5D in alcohol dependent patients: relationships among health-related quality of life, psychopathology and social functioning. Drug Alcohol Depend 2007;86:253–264.
15 Brogly S, Mercier C, Bruneau J, Palepu A, Franco E: Towards more effective public health programming for injection drug users: development and evaluation of the Injection Drug User Quality of Life Scale. Subst Use Misuse 2003;38:965–992.
16 Hubley A, Russell L, Palepu A: Injection Drug Use Quality of Life scale (IDUQOL): a validation study. Health Qual Life Outcomes 2005;3:43.
17 Morales-Manrique CC, Valderrama Z, Castellano-Gómez M, Aleixandre R, Palepu A: Cross cultural adaptation of the Injection Drug User Quality Of Life Scale (IDUQOL) in Spanish drug dependent population, with or without injectable consumption: Drug User Quality of Life Scale-Spanish (DUQOL-Spanish). Addict Behav 2007;32:1913–1921.
18 Lozano OM, Rojas AJ, Pérez C, Apraiz B, Sánchez F, Marín A: Test para la Evaluación de la Calidad de Vida en Adictos a Sustancias Psicoactivas (TECVASP): estudios de fiabilidad y validez. Trastornos Adict 2007;9:97–107.
19 Edwards G, Gross MM: Alcohol dependence: provisional description of a clinical syndrome. BMJ 1976;1:1058–1061.
20 Cella D, Chang CH: A discussion of item response theory and its applications in health status assessment. Med Care 2000;38(suppl II):II66–II72.
21 Hays RD, Morales LS, Reise SP: Item response theory and health outcomes measurement in the 21st century. Med Care 2000;38(suppl II):II28–II42.
22 Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C: Comparing alternative Rasch-based methods vs. raw scores in measuring change in health. Med Care 2004;42(suppl I):I25–I36.
23 Strong D, Lesieur H, Breen R, Stinchfield R, Lejuez CW: Using a Rasch model to examine the utility of the South Oaks Gambling Screen across clinical and community samples. Addict Behav 2004;29:465–481.
24 Rojas AJ, Lozano OM: Application of an IRT Polytomous Model for measuring health-related quality of life. Soc Indic Res 2005;74:369–394.
25 González-Saiz F, Lozano OM, Ballesta R, Silva T, Brugal MT, Bilbao I, Barrio G, Domingo-Salvany A, Bravo MJ, de la Fuente L: Validity of the Severity of Dependence Scale (SDS) construct applying the Item Response Theory to a non-clinical sample of heroin users. Subst Use Misus 2007;43:919–935.
26 APA, AERA, NCME, American Psychological Association, American Educational Research National Council on Measurement in Education: Standards for educational and psychological testing. Washington, 1999. http://www.apa.org/science/standards.html
27 Lozano OM, Rojas AJ, Pérez C, González-Saiz F, Ballesta R, Bilbao I: Validity evidences of the Health Related Quality of Life for Drug Abuser Test (HRQOLDA Test) based on Biaxial Model of Addiction. Psicothema 2008;20:311–317.
28 Wright BD, Mok MC: An overview of the family of Rasch measurement models; in Smith E, Smith R (ed): Introduction to Rasch Measurement. Chicago, Jampress, 2004.
29 Masters GN: A Rasch model for rating scales. Diss Abstr Int 1980;41:215A–216A.
30 Andrich D: Application of a psychometric rating model to ordered categories which are scored with successive integers. Appl Psychol Meas 1978;2:581–594.
31 Andrich D: A rating formulation for ordered response categories. Psychometrika 1978;43:561–573.
32 Wright BD, Masters GN: Rating Scale Analysis. Chicago, Mesa Press, 1982.
33 Rasch G: Probabilistic Models for Same Intelligence and Attainment Test. Chicago, University of Chicago Press, 1960.
34 Linacre JM: Winsteps (computer program and manual). Chicago, MESA Press, 2000.
35 Hambleton RK, Swaminathan H, Rogers HJ: Fundamentals of Item Response Theory. Sage, Newbury Park, 1991.
36 Wright BD, Linacre JM: Reasonable mean-square fit values. Rasch Measurement Trans 1994;8:370.
37 Lunz ME, Wright BD, Linacre JM: Measuring the impact of judge severity on examination scores. Appl Meas Educ 1990;3:331–345.
38 Linacre JM: What do infit and outfit, mean-square and standardized mean? Rasch Measurement Trans 2002;168:878.
39 Lai JS, Eton DT: Clinically meaningful gaps. Rasch Measurement Trans 2002;15:850.
40 Linacre JM: Comparing 'Partial Credit' and 'Rating Scale' Models. Rasch Measurement Trans 2000;14:768.
41 Andrich D, Jong JH, Sheridan BE: Diagnostic opportunities with the Rasch model for ordered response categories; in Rost J, Langeheine R (eds): Applications of Latent Trait and Latent Class Models in the Social Sciences. Münster, Waxmann, 1997, pp 59–70.
42 Ware JE, Bjorner JB, Kosinski M: Practical implications of Item Response Theory and Computerized Adaptive Testing. Med Care 2000;38(suppl II):II73–II82.