Modeling the World Health Organization Disability Assessment Schedule II using non-parametric item response models

FRANCISCA GALINDO-GARRE,¹ MARÍA DOLORES HIDALGO,² GEORGINA GUILERA,^{3,4} OSCAR PINO,⁵ J. EMILIO ROJO⁵ & JUANA GÓMEZ-BENITO^{3,4}

1 Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

2 Department of Behavioral Sciences Methods, University of Murcia, Murcia, Spain

3 Department of Behavioral Sciences Methods, Faculty of Psychology, University of Barcelona, Barcelona, Spain

4 Institute for Brain, Cognition and Behavior (IR3C), University of Barcelona, Barcelona, Spain

5 Department of Psychiatry, Hospital General Granollers – Benito Menni CASM, Granollers, Spain

---

Key words

WHO-DAS II, non-parametric IRT, Mokken scaling

Correspondence

Francisca Galindo-Garre,
Department of Epidemiology and
Biostatistics, VU University
Medical Center, Postbus 7057,
1007 MB Amsterdam, The
Netherlands.

Telephone (+31) 20 44444 90

Fax (+31) 20 44444 75

Email: f.galindo@vumc.nl

Received 18 October 2013;

revised 2 March 2014;

accepted 8 May 2014

Abstract

The World Health Organization Disability Assessment Schedule II (WHO-DAS II) is a multidimensional instrument developed for measuring disability. It comprises six domains (getting around, self-care, getting along with others, life activities and participation in society). The main purpose of this paper is the evaluation of the psychometric properties for each domain of the WHO-DAS II with parametric and non-parametric Item Response Theory (IRT) models. A secondary objective is to assess whether the WHO-DAS II items within each domain form a hierarchy of invariantly ordered severity indicators of disability.

A sample of 352 patients with a schizophrenia spectrum disorder is used in this study. The 36 items WHO-DAS II was administered during the consultation. Partial Credit and Mokken scale models are used to study the psychometric properties of the questionnaire.

The psychometric properties of the WHO-DAS II scale are satisfactory for all the domains. However, we identify a few items that do not discriminate satisfactorily between different levels of disability and cannot be invariantly ordered in the scale.

In conclusion the WHO-DAS II can be used to assess overall disability in patients with schizophrenia, but some domains are too general to assess functionality in these patients because they contain items that are not applicable to this pathology. Copyright © 2014 John Wiley & Sons, Ltd.

Introduction

Psychotic disorders are characterized by the high prevalence and severity of functional impairments (Green *et al.*, 2000; Harvey, 2010). The assessment of functional disability has become a key aspect for the evaluation of treatment and intervention-program effects, and, as a consequence, the interest on the development of standardized instruments to assess functional disability has increased. These instruments must have adequate psychometric features to achieve accurate measurements.

The World Health Organization (WHO) has developed a tool for assessing disability, called WHO Disability Assessment Schedule II (WHO-DAS II; WHO, 2000). This instrument contains questions to evaluate body, personal and social functioning. It has been translated and adapted to more than 10 languages and cultures, and at an international level numerous studies have shown the adequacy of its psychometric properties, both of the complete and reduced versions, in a variety of pathologies (see Federici and Meloni, 2010; Federici *et al.*, 2009). Psychiatric disorders are no exception, and the psychometric properties of the questionnaire have been analyzed for example in affective disorders (Chwastiak and Von Korff, 2003; Pösl *et al.*, 2007; Luciano *et al.*, 2010a, 2010b, 2010c), psychosis (Ulug *et al.*, 2001; Pyne *et al.*, 2003; Chopra *et al.*, 2004; McKibbin *et al.*, 2004), anxiety (Perini *et al.*, 2006), or schizophrenia (Guilera *et al.*, 2012).

Two approaches are commonly used to study the reliability and validity of these instruments: Classical test theory, which is concerned with the estimation of true scores and measurement errors and Item Response Theory (IRT), which is concerned with the responses to individual items. Compared to classical test theory, IRT models have several advantages:

- (a) invariant estimates of the psychometric properties of both items and subject characteristics are provided. In other words, the parameters that define an item do not depend upon the particular subject sample used, and the parameters that define the subject do not depend upon the particular item sample used;
- (b) a detailed description of the behavior of individual items across all levels of the trait is provided;
- (c) they enable precision (reliability) indices to be derived for both the test and the items, and more importantly, precision can be evaluated for each level of the measured trait;
- (d) the theory makes it easier to develop scales and tests that are adapted to subjects, for a given level of the trait being measured and for the whole of its continuum.

Within the IRT modeling approach, an alternative scaling methodology that is becoming popular in health research (Watson *et al.*, 2008, 2011) is Mokken scaling (Mokken, 1971; Molenaar, 1997). This methodology uses non-parametric models for assessing the relationship between items and was developed from the Guttman method to investigate hierarchies of items within scales. Non-parametric IRT models are based on less-restrictive assumptions than parametric IRT models, and they provide person and item parameters that are intuitive and easy to interpret. Establishing the hierarchical properties of items within a scale provides additional information when working with its total score. A hierarchical scale has items that can be ordered according to their mean scores in the total group, and the ordering is invariant across individuals.

Despite its promising utility, none of the studies published so far in schizophrenia have used the IRT framework to investigate the psychometric properties of the WHO-DAS II in the complete form (36 items). Furthermore, there is little information on the effectiveness of each item and their response options in discriminating between different disability levels, among patients with this disorder.

The purpose of the present study is to provide a contribution to the clinical utility of the 36-item WHO-DAS II exploring its metric quality in a sample of patients with a diagnosis of schizophrenia from the perspective of IRT using non-parametric models.

Methods

Description of the sample

To be included in the study patients were required to be between 18 and 55 years old, meet the diagnostic criteria of either schizophrenia, schizoaffective disorder, or schizophreniform disorder (DSM-IV-TR; American Psychiatric Association, 2000), and give written consent to participate. Patients were excluded if they were suffering from a severe medical or neurological condition, or another psychiatric disorder that required treatment. Patients were also excluded if they were participating in a clinical trial, suffering an episode of major depression, or clearly exhibiting difficulties with reading or writing.

The sample included a total of 352 patients with a schizophrenia spectrum disorder, 69.3% male and 30.7% female, with a mean age of 36.70 (standard deviation [SD] = 8.27) and a range between 18 and 55 years old. Among them 309 were diagnosed with schizophrenia (87.8%), 35 with schizoaffective disorder (9.9%) and eight with schizophreniform disorder (2.3%). The average duration of illness was 140.65 (SD = 101.50) months, and the

average number of prior hospital admissions was 0.14 (SD = 0.40). Table 1 lists the main socio-demographic characteristics of the sample.

Instruments and procedure

The WHO-DAS II (WHO, 2000) is a multidimensional instrument developed for measuring disability. There is a self-administered and an interviewer administered version, each of which are available in long and short (screening) form. The complete version of the instrument contains 36 items on functioning and disability with a recall period of 30 days, and comprises six domains: D1 "Understanding and communicating" (six items), D2 "Getting around" (five items), D3 "Self-care" (four items), D4 "Getting along with others" (five items), D5 "Life activities" (household and school or work; eight items), and D6 "Participation in society" (eight items). Items

can be scored on a five-point scale ranging from 1 = none to 5 = extreme/cannot do. For all domains, the WHO-DAS II provides different profiles and a single disability summary score that can be obtained by using the SPSS syntax offered by the WHO. Scores may vary from 0 to 100 where higher scores reflect greater disability. Participants who are not currently working answer 32 items, while those that are working or studying respond to the entire group of items.

The Ethics Committee of the University of Barcelona approved this multi-center study. Patients were recruited through 40 outpatient psychiatric clinics across Spain, after having given their written consent to participate. They completed a comprehensive assessment battery including measures on psychopathology, functionality and quality of life.

Statistical methods

First, means, standard deviations and frequency distributions were calculated for all WHO-DAS II items. Then principal component analyses with Varimax rotation were performed to study the dimensionality of the different WHO-DAS II scales; moreover, a classical item analysis was realized. Analyses were performed with the R package *psych* (Revelle, 2013).

The R package *Mokken* (Van der Ark, 2007) was used to study the hierarchy of the different scales of the WHO-DAS II. First, the Monotone Homogeneity Model (MHM) was fitted to determine whether item scores are ordered on the latent trait. This model assesses three assumptions: (i) unidimensionality, (ii) monotonicity, and (iii) conditional independence. These assumptions can be tested by checking the restrictions on the scalability coefficients H (Sijtsma and Molenaar, 2002, theorem 4.3), i. e. the total H coefficient value, the H coefficient for each item, and the H coefficient for each pair of items. The scalability coefficients H must be between zero and one. The procedures to check these restrictions are the automated item selection procedure (Molenaar and Sijtsma, 2000) and the item rest score regression. The scalability coefficient H (Mokken, 1971) was computed to determine the strength of the relationship of each item j with the latent trait. A set of items form a scale if the H coefficient for each pair of items is higher or equal than 0.3. Furthermore, scales are classified according to the following criteria for the H value: (a) $0.3 \leq H < 0.4$: weak scale, (b) $0.4 \leq H < 0.5$: medium scale, and (c) $H \geq 0.5$: strong scale. The reliability of the scale was checked with the Cronbach's alpha and the Molenaar Sijtsma (MS) statistic, which is a more accurate reliability coefficient. For a

Table 1. Socio-demographic characteristics of the sample

Variable	<i>n</i>	%
Marital status ^a		
Single	294	83.5
Married	25	7.1
Living with couple	11	3.1
Widow/er	2	0.6
Separated or divorced	19	5.4
Living arrangement		
Original family	240	68.2
Own family	50	14.2
Friends	5	1.4
Sheltered housing	12	3.4
Alone	37	10.5
Other	8	2.3
Educational level ^a		
Functional illiterate	7	2.0
Primary education	121	34.4
Secondary education	168	47.7
University education	55	15.6
Employment status		
Student	15	4.3
Wage earner	74	21.0
Self-employed	13	3.7
Non-paid work	3	0.9
Disability	136	38.6
Unemployed	53	15.1
Retired	29	8.2
Housewife	9	2.6
Other	20	5.7

^aDue to missing values, percentages do not sum 100

description of the properties of these coefficients, see Van der Ark (2012).

Then, the Double Monotonicity Model (DMM) for polytomous items was fitted. This model yields a good fit when the previously described assumptions hold and when the items are ordered among subjects. This property, called invariant item ordering (IIO), is a necessary condition for a scale to be hierarchical, and it can be tested by the method of manifest IIO (MIIO; Ligetvoet *et al.*, 2010). Items involved in violations of the IIO assumption are removed from the questionnaire by the backward method (Van der Ark, 2012). After IIO was established, the H^T coefficient was calculated to assess the precision of the item ordering (Ligetvoet *et al.*, 2010). The H^T coefficient is evaluated following the criteria described for the H coefficient.

Finally, a more restrictive parametric IRT model, the Partial Credit Model (PCM) (Embretson and Reise, 2000), was fitted for each scale with the R package eRm (Mair and Hatzinger, 2007). This model describes the probability that a person responds into a category conditional on the location of the person on the continuum of functional disabilities for each scale. The model has four difficulty parameters for each item, which is the number of response categories minus one. For a polytomous item, the difficulty parameters can be considered as step difficulties associated with the transition from one category to the next. The fit of the PCM was assessed by mean squared INFIT and OUTFIT statistics. Values higher than 1.3 or lower than 0.7 identify items that need to be revised (Smith *et al.*, 2008). Person item maps are used to illustrate the range and position of the item and person measure distributions because these plots orders the ability levels of the patients and the difficulty levels of the items.

Results

Descriptive item analysis

Table 2 reports descriptive statistics for all the scales of the WHO-DAS II. From Table 2 we can conclude that schizophrenia patients do not have many difficulties to perform the items of any of the scale. There were floor effects in many items of all the scales because a high proportion of the patient report the first category none for most of the items. For the D1 “Understanding and communicating” scale, the categories 4 and 5 have a proportion of response lower than 0.05 for all the items. For the D2 “Getting around” and D3 “Self care”, we found the same low proportion for the response categories 3, 4 and 5. The proportion of patients having difficulties with the items is slightly higher for the scales D5 “Life activities” and D6 “Participation in society”.

Principal component analysis

Table 3 illustrates that all WHO-DAS II scales are unidimensional. The percentage of explained variance ranges between 47% for the getting around scale and 63% for the life activities scales. All factor loadings are higher than 0.35. Since all the scales are one-dimensional, we first fit non-parametric IRT models to study the relationships between items.

Mokken scaling

Mokken scaling analysis is performed for each WHO-DAS II scale apart in order to assess their scaling properties and to investigate whether there is an invariant item ordering between the items in the scales.

The Monotone Homogeneity Model (MHM)

Table 4 shows the reliability and the scalability coefficient for each WHO-DAS II scale. The scales D1 “Understanding and communicating”, D4 “Getting along with people”, D5 “Life activities”, and D6 “Participation in society” are reliable scales ($MS > 0.8$, $\alpha > 0.8$); the H coefficients are higher than 0.5 indicating a strong relationship between their items and the measured latent trait. The reliability coefficient for the D2 “Getting around” and D3 “Self care” scales are lower than 0.7, and the scalability coefficients H show medium relationships between the items and the latent trait. Table 5 illustrates that within the D2 “Getting around” scale the items D23 *Moving around inside your home* and D24 *Getting out of your home* achieve an H coefficient that does not meet the scalability criteria ($H \geq 0.3$). From the low mean scores (1.02 and 1.12 respectively) we can conclude that there is a floor effect for these items. The percentage of patient reporting no difficulties for item D23 was 98%, and for item D24 was 88%. These items cannot discriminate between patients from this population. Further, there are no violations of the monotony assumption for any of the scales. We can conclude that the MHM fits except for the D2 “Getting around” scale.

The Double Monotonicity Model (DMM)

To fit the DMM, the ordering of the items is evaluated for each scale. The MIIO method is used to find violations of the item invariant ordering assumption. We do not find any violation of this assumption. However, the H^T coefficients for the D1 “Understanding and communication”, D5 “Life activities” and D6 “Participation in society” scales do not meet the criteria ($H^T \geq 0.3$), which indicated that the achieved ordering of the items is not precise. The D2

Table 2. Mean, standard deviation (SD) and proportion for each level of response

Scale	Item	Mean	SD	1	2	3	4	5
Understanding and communicating	D1.1	1.82	1.01	0.49	0.28	0.15	0.04	0.02
	D1.2	1.69	0.70	0.52	0.31	0.13	0.03	0.00
	D1.3	1.79	0.92	0.52	0.25	0.17	0.07	0.00
	D1.4	1.54	0.66	0.63	0.24	0.10	0.03	0.00
	D1.5	1.43	0.25	0.57	0.43	0.00	0.00	0.00
	D1.6	1.55	0.39	0.52	0.42	0.07	0.00	0.00
Getting around	D2.1	1.29	0.53	0.81	0.13	0.02	0.02	0.01
	D2.2	1.06	0.08	0.96	0.03	0.01	0.00	0.00
	D2.3	1.02	0.02	0.98	0.02	0.00	0.00	0.00
	D2.4	1.12	0.11	0.88	0.12	0.00	0.00	0.00
	D2.5	1.18	0.26	0.87	0.10	0.02	0.01	0.00
Self care	D3.1	1.06	0.05	0.94	0.06	0.00	0.00	0.00
	D3.2	1.01	0.01	0.99	0.01	0.00	0.00	0.00
	D3.3	1.04	0.04	0.96	0.04	0.00	0.00	0.00
	D3.4	1.21	0.22	0.81	0.17	0.02	0.00	0.00
Getting along with people	D4.1	1.78	0.43	0.35	0.53	0.12	0.00	0.00
	D4.2	1.53	0.34	0.52	0.44	0.04	0.00	0.00
	D4.3	1.34	0.29	0.70	0.27	0.03	0.00	0.00
	D4.4	2.37	1.40	0.33	0.20	0.27	0.18	0.02
	D4.5	1.85	0.58	0.37	0.40	0.22	0.00	0.00
Life activities	D5.2	1.33	0.27	0.70	0.28	0.02	0.00	0.00
	D5.3	1.35	0.25	0.66	0.33	0.01	0.00	0.00
	D5.4	1.48	0.52	0.64	0.25	0.10	0.01	0.00
	D5.5	1.44	0.27	0.57	0.42	0.01	0.00	0.00
	D5.8	1.45	0.34	0.60	0.36	0.04	0.00	0.00
	D5.9	1.55	0.77	0.64	0.22	0.09	0.03	0.01
	D5.10	1.58	0.86	0.63	0.22	0.10	0.02	0.02
Participation in society	D5.11	1.76	0.95	0.52	0.28	0.15	0.03	0.02
	D6.1	1.58	0.38	0.48	0.45	0.07	0.00	0.00
	D6.2	1.38	0.49	0.72	0.20	0.06	0.02	0.00
	D6.3	1.28	0.20	0.72	0.28	0.00	0.00	0.00
	D6.4	1.80	0.78	0.45	0.35	0.17	0.02	0.01
	D6.5	1.84	0.97	0.48	0.27	0.18	0.06	0.01
	D6.6	1.43	0.32	0.61	0.36	0.03	0.00	0.00
	D6.7	1.82	0.97	0.51	0.25	0.17	0.08	0.00
	D6.8	1.54	0.39	0.53	0.40	0.07	0.00	0.00

“Getting around” scale has a coefficient of 0.39 that suggests weak support for the IIO assumption, and the scales D3 “Self care” and D4 “Getting along with people” achieve H^T coefficients of 0.63 and 0.55 respectively suggesting a strong support for the IIO assumption. Our results suggest that the scale D2 “Getting around” should be revised, particularly the items D23 and D24.

Partial Credit Model

Finally, we fit a PCM to compare the results achieved by non-parametric models to the results achieved by more

restricted parametric IRT models. Both mean squared INFIT and OUTFIT statistics in Table 5 show that all scales contain items that achieve statistics outside the permissible range (0.7–1.3). Person item maps in Figure 1 compare the range and position of item measure distributions to the range and position of latent trait distributions. Items should be located at each point on the scale and items should cover all areas of ability. Figure 1 shows that the maps for the D2 “Getting around” and D3 “Self care” scales are remarkably bad because of the ability range assessed and the position of the items across the range. We can conclude that the PCM is too restrictive and

Table 3. Factor loadings and percentages of explained variance for each scale of the WHO-DAS II

Scale	Item	Item content	Factor loadings
Understanding and communicating	D1.1	Concentrating on doing something for 10 minutes	0.69
	D1.2	Remembering to do important things	0.79
	D1.3	Analyzing and finding solutions to problems	0.86
	D1.4	Learning a new task	0.77
	D1.5	Generally understanding	0.69
	D1.6	Starting and maintaining a conversation	0.74
Explained variance			55%
Getting around	D2.1	Standing for long periods 30 minutes	0.77
	D2.2	Standing up	0.85
	D2.3	Moving around inside your home	0.38
	D2.4	Getting out of your home	0.45
	D2.5	Walking a long distance such as a kilometer	0.81
Explained variance			47%
Self care	D3.1	Washing whole body	0.79
	D3.2	Dressed	0.78
	D3.3	Eating	0.76
	D3.4	Staying by yourself for a few days	0.48
Explained variance			51%
Getting along with people	D4.1	Dealing with people you do not know	0.84
	D4.2	Maintaining a friendship	0.81
	D4.3	Getting along with people who are close to you	0.66
	D4.4	Making new friends	0.90
	D4.5	Sexual activities	0.65
Explained variance			61%
Life activities	D5.2	Taking care of your household responsibilities	0.66
	D5.3	Doing most important household tasks well	0.68
	D5.4	Getting all the household work done	0.79
	D5.5	Getting your household work quickly as needed	0.75
	D5.8	Your day-to-day work/school	0.80
	D5.9	Doing your most important work tasks well	0.86
	D5.10	Getting all the work done that you need to do	0.89
	D5.11	Getting your work done as quickly as needed	0.89
Explained variance			63%
Participation in society	D6.1	Problem in joining in community activities	0.68
	D6.2	Barriers or hindrances in the world around	0.72
	D6.3	Living with dignity	0.73
	D6.4	Time spend on your health condition	0.77
	D6.5	Emotionally affected by health condition	0.81
	D6.6	A drain on the financial resources	0.75
	D6.7	How much of a problem did your family have	0.77
	D6.8	Problem doing things by yourself for relaxation or pleasure	0.74
Explained variance			56%

cannot be used to assess the psychometric properties of this scale.

Discussion

The present study is the first that uses parametric and non-parametric IRT models to assess the psychometric

properties of the Spanish version of the WHO-DAS II in their complete form with a large sample of patients diagnosed with schizophrenia.

First, floor effects were detected for several items across different domains, especially those related to the dimension D2 “Getting around” and D3 “Self care”. This result is not surprising because of the difficulty that patients with

Table 4. Scalability coefficient (H), standard error (SE), reliability rho, Cronbach alpha, and H^T for each scale of the WHO-DAS II

Scale	H	SE	Reliability MS	Cronbach alpha	H^T
Understanding and communicating	0.53	0.06	0.84	0.83	0.12
Getting around	0.48	0.10	0.75	0.65	0.39
Self care	0.46	0.19	0.61	0.50	0.63
Getting along with people	0.63	0.05	0.85	0.81	0.55
Life activities	0.69	0.06	0.93	0.91	0.19
Participation in society	0.58	0.05	0.90	0.87	0.26

a serious mental illness, such as schizophrenia, have to express their experience (Doyle *et al.*, 1999; Bowie *et al.*, 2007). These authors postulate that the lack of insight and cognitive impairments can alter the perception that patients have of their illness. Therefore, self-report instruments may underestimate the level of disability. Guilera *et al.* (2012) show that since 60% of patients achieved the best score (i.e. zero score) in the subscales D2 and D3, these two scales present a floor effect and they are not good candidates to discriminate between schizophrenia patients with low disability levels. The lack of awareness of the disease could be one of the aspects that influence self-perception and therefore the answers to some of the items.

Second, each of the six WHO-DAS II scales were unidimensional, and the factor loadings achieved were higher than 0.35. These results support evidence that the item responses for each dimension are reasonably dominated by a single latent trait. The unidimensionality of the 12 items screening version was confirmed by Sousa *et al.* (2010) for dementia patients and by Luciano *et al.* (2010c) for patients with depression. The present study adds evidence to the ongoing process of validity of the WHO-DAS II by supporting also the unidimensionality of the six domains of the complete 36 items version.

Third, regarding the results from non-parametric IRT, the MHM achieves an excellent fit for the scales D1 “Understanding and communicating”, D4 “Getting along with people”, D5 “Life activities” and D6 “Participation in society”. The fit of the model is acceptable for the scales D3 “Getting around” and D2 “Self-care”. A good fit of the model means that the mean of the latent trait increases as the sum score increases. In most of the practical situations, the sum score can be used to order patients stochastically on the trait (Van der Ark, 2005). In general, the WHO-

DAS II items discriminate well between different levels of disability. This implies that subjects with more difficulties in functioning have higher expected scores for the items in the scale. However, the MHM did not fit well for two items included in the D3 scale, in particular the D23 *Moving around inside your home* and D24 *Getting out of your home* items show floor effects because they assessed competences that have little impact on the functioning for schizophrenic patients. This result should be taken into account to interpret the total scores obtained by patient with the same pathology as the patients assessed in this study.

The DMM did not fit for D1 “Getting around”, D2 “Self-care”, D5 “Life activities” and D6 “Participation in society”. For these scales, the items cannot be invariantly ordered. This means that patients with the same level in the latent trait may have difficulties with different items. In other words, the item response category characteristic curves cross between items. The DMM achieved a good fit for the D3 “Getting around” and D4 “Getting along with people”. This means that we can order activities related with personal care and the ability to establish personal relationships with other people. These results are in line with the findings of Sousa *et al.* (2010) with the 12 items screening form in dementia patients; in testing hierarchical structure, they found that items meet MHM assumption but not DMM because of the two items of the D1 “Getting around” domain.

Finally, the more restrictive PCM did not fit to any scale. This model was too restrictive because it implies an order not only in the subject ability and in the items, but also in the item response categories. The later assumption may be excessively restrictive because the distribution of the items scores shows a positive skewness. Lower scores are more commonly chosen which would be expected if the PCM fits.

Table 5. Scalability coefficient (H) with the standard error (SE) and Mean squared INFIT statistic (INFIT MSQ) and Mean squared OUTFIT statistic (OUTFIT MSQ) for each item per scale of the WHO-DAS II

Scale	Item	H	SE	OUTFIT MSQ ^a	INFIT MSQ ^a
Understanding and communicating	D1.1	0.46	0.08	1.05	1.07
	D1.2	0.56	0.07	0.71	0.74
	D1.3	0.60	0.06	<i>0.60</i>	<i>0.60</i>
	D1.4	0.56	0.07	0.78	0.78
	D1.5	0.52	0.08	0.96	0.96
	D1.6	0.42	0.09	1.12	1.11
Getting around	D2.1	0.44	0.13	0.69	0.71
	D2.2	0.66	0.09	<i>0.23</i>	<i>0.39</i>
	D2.3	<i>0.24</i>	0.15	0.89	1.10
	D2.4	<i>0.25</i>	0.13	<i>1.50</i>	1.30
	D2.5	0.52	0.12	<i>0.58</i>	<i>0.62</i>
Self care	D3.1	0.53	0.19	<i>0.61</i>	<i>0.67</i>
	D3.2	0.73	0.01	<i>0.10</i>	<i>0.31</i>
	D3.3	0.39	0.24	0.92	0.84
	D3.4	0.38	0.17	0.83	0.78
Getting along with people	D4.1	0.70	0.06	<i>0.68</i>	<i>0.69</i>
	D4.2	0.68	0.06	0.74	0.73
	D4.3	0.55	0.08	0.90	1.00
	D4.4	0.69	0.05	<i>0.49</i>	<i>0.50</i>
	D4.5	0.50	0.08	<i>1.37</i>	1.30
Life activities	D5.2	0.55	0.11	1.27	<i>1.31</i>
	D5.3	0.56	0.09	1.02	1.23
	D5.4	0.64	0.09	1.18	1.07
	D5.5	0.67	0.08	0.98	1.02
	D5.8	0.70	0.08	0.81	0.83
	D5.9	0.73	0.05	0.70	<i>0.64</i>
	D5.10	0.76	0.05	<i>0.52</i>	<i>0.49</i>
Participation in society	D5.11	0.77	0.05	<i>0.69</i>	<i>0.63</i>
	D6.1	0.53	0.06	1.13	1.09
	D6.2	0.57	0.07	<i>0.55</i>	0.81
	D6.3	0.61	0.08	0.69	0.82
	D6.4	0.60	0.05	0.91	0.84
	D6.5	0.59	0.05	0.80	0.84
	D6.6	0.61	0.07	0.86	0.85
	D6.7	0.56	0.07	1.04	1.05
	D6.8	0.57	0.06	0.83	0.89

^aValues higher than 1.3 or lower than 0.7 are printed in italics to identify items that need to be revised.

In summary, the current study provides new evidence on the usefulness and measurement quality of the WHO-DAS II, from the IRT framework, complementing the findings of Guilera *et al.* (2012) from the classical test theory. The results lead to the conclusion that some subscales of the WHO-DAS II may be too general to assess functionality in patients with schizophrenia because they contain some items that are not applicable to this pathology. This study suggests that a reduced instrument should be elaborated with the items that discriminate between different levels of disability for this group. The 12 items

screening version of the WHO-DAS II does not solve the problem (see Sousa *et al.*, 2010) since it also contains items that are not applicable to the analyzed pathology. In this sense, the utility of Mokken scaling in terms of producing shorter scales is indubitable, and the present study can help to identify the specific areas of disability in schizophrenic patients. The presence of non-discriminative items is a common problem shared by generic tools to assess disability or quality of life. This problem raises the controversy over the application of generic or specific instruments, or whether to use of both can

Figure 1 Person item map for the Partial Credit Model

complete the information about a subject's position on the trait.

Acknowledgements

This work was supported by Spain's Ministry of Economy and Competitiveness (grant number PSIC2012-32275); the

Agency for the Management of University and Research Grants from the Government of Catalonia (grant number 2009SGR00822); and by Pfizer, S.L.U.

Declaration of interest statement

The authors have no competing interests.

References

- American Psychiatric Association. (2000) *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV TR)*, Washington, DC, American Psychiatric Association.
- Bowie C.R., Twamley E.W., Anderson H., Halpern B., Patterson T.L., Harvey P.D. (2007) Self-assessment of functional status in schizophrenia. *Journal of Psychiatric Research*, 41(12), 1012–1018.
- Chopra P.K., Couper J.W., Herman H. (2004) The assessment of patients with long-term psychotic disorders: application of the WHO Disability Assessment Schedule II. *Australian and New Zealand Journal of Psychiatry*, 38(9), 753–759.

- Chwastiak L.A., Von Korff M. (2003) Disability in depression and back pain: evaluation of the World Health Organization Disability Assessment Schedule (WHO DAS II) in a primary care setting. *Journal of Clinical Epidemiology*, **56**(6), 507–514.
- Doyle M., Flanagan S., Browne S., Clarke M., Lydon D., Larkin E., O'Callaghan C. (1999) Subjective and external assessments of quality of life in schizophrenia: relationship to insight. *Acta Psychiatrica Scandinavica*, **99**(6), 466–472.
- Embretson S.E., Reise S.P. (2000) *Item Response Theory for Psychologists*, Mahwah, NJ, Lawrence Erlbaum Associates.
- Federici S., Meloni F. (2010) WHODAS II: disability self-evaluation in the ICF conceptual frame. In *International Encyclopedia of Rehabilitation* Stone J., Blouin M. (eds), Buffalo, NY, State University of New York, 1–22.
- Federici S., Meloni F., Lo Presti A. (2009) International literature review on WHODAS II (World Health Organization Disability Assessment Schedule II). *Life Span and Disability*, **12**, 83–110.
- Green M.F., Kern R.S., Braff D.L., Mintz J. (2000) Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the 'right stuff? *Schizophrenia Bulletin*, **26** (1), 119–136.
- Guilera G., Gómez-Benito J., Pino O., Rojo J.E., Cuesta M.J., Martínez-Arán A., Safont G., Tabarés-Seisdedos R., Vieta E., Bernardo M., Crespo-Facorro B., Franco M., Rejas J. (2012) Utility of the World Health Organization Disability Assessment Schedule II in schizophrenia. *Schizophrenia Research*, **138** (2), 240–247.
- Harvey P.D. (2010) Cognitive functioning and disability in schizophrenia. *Current Directions in Psychological Science*, **19**(4), 249–254.
- Ligtvoet R., van der Ark L.A., Bergsma W., Sijtsma K. (2010) Investigating an invariant item ordering for polytomous scored items. *Educational and Psychological Measurement*, **70**(4), 578–595.
- Luciano J.V., Ayuso-Mateos J.L., Aguado J., Fernandez A., Serrano-Blanco A., Roca M., Haro J.M. (2010a) The 12-item World Health Organization Disability Assessment Schedule II (WHO-DAS II): a nonparametric item response analysis. *BMC Medical Research Methodology*, **10**(1), 45.
- Luciano J.V., Ayuso-Mateos J.L., Fernandez A., Aguado J., Serrano-Blanco A., Roca M., Haro J.M. (2010b) Utility of the twelve-item World Health Organization Disability Assessment Schedule II (WHO-DAS II) for discriminating depression "caseness" and severity in Spanish primary care patients. *Quality of Life Research*, **19**(1), 97–101.
- Luciano J.V., Ayuso-Mateos J.L., Fernandez A., Serrano-Blanco A., Roca M., Haro J.M. (2010c) Psychometric properties of the twelve-item World Health Organization Disability Assessment Schedule II (WHO-DAS II) in Spanish primary care patients with a first major depressive episode. *Journal of Affective Disorders*, **121**(1), 52–58.
- Mair P., Hatzinger R. (2007) Extended Rasch Modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, **20**(9), 1–20.
- McKibbin C., Patterson T.L., Jeste D.V. (2004) Assessing disability in older patients with schizophrenia: results from the WHODAS-II. *Journal of Nervous and Mental Disease*, **192** (6), 405–413.
- Mokken R. (1971) *A Theory and Procedure of Scale Analysis with Applications in Political Research*, New York, De Gruyter.
- Molenaar I.W. (1997) Nonparametric models for polytomous responses. In *Handbook of Modern Item Response Theory* van der Linden W., Hambleton K.K. (eds), pp. 367–380, New York, Springer-Verlag.
- Molenaar I.W., Sijtsma K. (2000) *MSP5 for Windows*, Groningen, iec ProGAMMA.
- Perini S.J., Slade T., Andrews G. (2006) Generic effectiveness measures: sensitivity to symptom change in anxiety disorders. *Journal of Affective Disorders*, **90**(2), 123–130.
- Pösl M., Cieza A., Stucki G. (2007) Psychometric properties of the WHODAS II in rehabilitation patients. *Quality of Life Research*, **16**(9), 1521–1531.
- Pyne J.M., Sullivan G., Kaplan R., Williams D.K. (2003) Comparing the sensitivity of generic effectiveness measures with symptom improvement in persons with schizophrenia. *Medical Care*, **41**(2), 208–227.
- Revelle W. (2013) Psych. R package version 1.3.2. <http://CRAN.R-project.org/package=psych> [14 January 2014].
- Sijtsma K., Molenaar I. (2002) *Introduction to Non-parametric Item Response Theory*, Thousand Oaks, CA, Sage.
- Smith A.B., Rush R., Fallowfield L.J., Velikova G., Sharpe M. (2008) Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, **8**(1), 1–33.
- Sousa R.M., Dewey M.E., Acosta D., Jotheeswaran A.T., Castro-Costa E., Ferri C.P., Guerra M., Huang Y., Jacob K.S., Rodriguez Pichardo J. G., García Ramírez N., Llibre Rodríguez J., Calvo Rodríguez M., Salas A., Sosa A.L., Williams J., Prince J. (2010) Measuring disability across cultures – the psychometric properties of the WHODAS II in older people from seven low- and middle-income countries. The 10/66 Dementia Research Group population – based survey. *International Journal of Methods in Psychiatric Research*, **19**(1), 1–17.
- Ulug B., Ertugrul A., Gögüs A., Kabakçi E. (2001) Yetiyitimi Değerlendirme Çizelgesinin (WHODAS II) sızofreni hastalarında geçerlilik ve güvenilirliği [Reliability and validity of the Turkish version of the World Health Organization Disability Assessment Schedule-II (WHODAS II) in schizophrenia]. *Türk Psikiyatri Dergisi*, **12**(2), 121–130.
- Van der Ark L.A. (2005) Stochastic ordering of the latent trait by the sum score under various polytomous IRT models. *Psychometrika*, **70** (2), 283–304.
- Van der Ark L.A. (2007) Mokken scale analysis in R. *Journal of Statistical Software*, **20**(11), 1–19.
- Van der Ark L.A. (2012) New developments in Mokken scale analysis. *Journal of Statistical Software*, **48**(5), 1–27.
- Watson R., Deary I.J., Shipley B. (2008) A hierarchy of distress: Mokken scaling of the GHQ-30. *Psychological Medicine*, **38**(4), 575–579.
- Watson R., van der Ark A., Lin L., Fieo R., Deary I., Meijer R. (2011) Item response theory: how Mokken Scaling can be used in clinical practice. *Journal of Clinical Nursing*, **21**(19–20), 2736–2749.
- World Health Organization (WHO). (2000) *World Health Organization: Disability Assessment Schedule II (WHO-DAS II)*, Geneva, WHO.