Validity and Reliability of the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) in Turkish Psychiatry Patients and Healthy Controls

Duygu ASLAN KUNT[1], Ferhan DEREBOY[2]

SUMMARY

Objective: The present study aimed to investigate psychometric properties of a series of disability scores obtained from Turkish version of the WHODAS 2.0 interviewer-, self- and proxy-administered forms consisting of either 36 items or 12 items.

Methods: Following the translation, 35 patients with a psychiatric diagnosis and 35 healthy controls between 18 and 65 years of age self-rated their functional impairment on the WHODAS. In addition, each participant was rated by a relative and by one or two clinicians on the pertinent WHODAS forms. In order to collect evidence for validity and reliability of WHODAS general disability and domain scores, we employed a series of Student's t-tests, ROC analyses, logistic regression analyses, intraclass and Pearson's correlation analyses, Cronbach's alpha and item-total statistics.

Results: Regarding general disability scores, in both clinical sample and healthy controls, all three types of 36-item WHODAS displayed satisfactory or higher validity and reliability coefficients. On the other hand, for 12-item version, only the interviewer-rated form demonstrated satisfactory results only in the clinical sample. Domain disability scores yielded by the 36-item forms were generally associated with adequate or acceptable coefficients in the clinical sample, while the coefficients were unacceptable in the control group.

Conclusion: The 36-item WHODAS interviewer-, proxy- and self-rated forms are suitable to assess general disability in Turkish mental health consumers and in healthy subjects. Among the 12-item WHODAS forms, the interviewer-rated form emerges as the sole instrument with comparable validity and reliability for measuring general disability in psychiatric patients. The domain disability scores derived from the long form and general disability scores derived from the short form is suitable for evaluating clinical subjects, but not healthy subjects.

Keywords: WHODAS 2.0, disability, reliability, validity

INTRODUCTION

Disability rendered by medical or mental disorders involves impaired basic life activities with important legal and clinical implications. Understanding the association of the disease with the disability is imperative for recovery of functional status and quality of life of affected individuals (Druss et al. 2000). Disability pertains to several domains of living including one's activities and social relations in addition to functionality and physical integrity of one's body (Üstün et al. 2010). There is no consensus on the life domains which should be measured by the disability scales. There are also remarkable differences among the existing instruments in terms of adopted terminology, theoretical framework, and evaluation strategy (Sheehan et al. 1996).

During development of the Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (DSM-5), the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) was decided to replace the Global Assessment of Functioning Scale (GAFS) of the previous DSM versions to assess impaired general functioning and disability. The

Received: 06.02.2017 - **Accepted:** 24.07.2017

[1]Psychiatrist, State Hospital, Department of Psychiatry, Aydın, [2]Prof., Faculy of Medicine, Adnan Menderes University, Department of Psychiatry, Aydın, Turkey.

e-mail: *drduyguaslan@gmail.com*

https://www.doi.org/10.5080/u20481

WHODAS 2.0 has been designed to evaluate the extent of impairment in affected individual's activity and participation in social life regardless of medical diagnosis, hence requires consideration of one's environmental conditions (Üstün et al. 2010). WHODAS 2.0 is a tool that assesses medical and psychiatric disorders while including environmental effects. An important aspect of the GAFS (American Psychiatric Association 2000) was its exclusive focus on impairment in psychological, social and occupational functioning, and ignorance of impairment in physical functioning, which is also addressed by WHODAS 2.0

The World Health Organization Disability Assessment Schedule II (WHODAS-II), an interviewer-rated measure, was previously validated in Turkish by Uluğ et al. (2001) in a sample consisting of patients with a diagnosis of schizophrenia. Current version the instrument referred to as WHODAS 2.0, however, entails self- and proxy-administered forms in addition to interviewer-administered form. The psychometric properties of the general disability and domain disability scores obtained from these three forms were not investigated in Turkey previously. The goal of the present study is to translate and validate the 36-item and 12-item WHODAS 2.0 versions, each available in interviewer-, proxy-, and self-administered forms, using data obtained from psychiatric patients and healthy controls. In this study, we aimed to collect pieces of validity evidence through testing the following five validity hypotheses, the last two of which being also relevant for reliability evaluation (Sireci and Sukin 2013).

1. The disability scores of the clinical sample are to be found significantly higher than those of the control sample and the difference between the two samples is to be large (Cohen's d ≥ 0.80) in each comparison (Cohen 1988).
2. The general disability scores used with optimal cutoffs are to predict subjects' group membership with modest (R^2 = 0.13 - 0.25) or large (R^2 ≥ 0.26) accuracy, and are to differentiate clinical subjects from control subjects with modest (Diagnostic Odds Ratio - DOR = 4.0-6.9) or large (DOR ≥ 7.0) accuracy (Glass et al. 2003, Chen et al 2010)
3. The level of agreement among general disability and domain disability scores yielded by different administration modes of the full or short versions are to be acceptable (ICC = 0.40-0.59) or sufficient (ICC≥0.60) levels. As well, the level of agreement between the general disability scores yielded by identical administration modes of the full and short versions are to be modest or sufficient considering the same ICC benchmarks for psychometric investigations (Kraemer et al. 2012).
4. Internal consistency of the WHODAS 2.0 scales and subscales are to be estimated at acceptable (Cronbach's alpha = 0.70 - 0.79) or sufficient (alpha ≥ 0.80 for measuring samples, alpha ≥ 0.90 for evaluating individuals) levels (Nunnally and Bernstein 1994).
5. The level of agreement between general disability or domain disability scores obtained from ratings by two interviewers are to be estimated at acceptable/questionable (ICC = 0.40-0.59) or sufficient (ICC ≥ 0.60) levels. (Kraemer et al. 2012, Clarke et al. 2013).

METHOD

Translation Study

Following a written permission from the World Health Organization (WHO), we translated all WHODAS 2.0 forms into Turkish in line with the "process of translation and adaptation of instruments" elucidated in the WHO website (World Health Organization 2015). First, the forms were translated into Turkish by the first author considering conceptual and cultural equivalence instead of linguistic equivalence. The first Turkish draft was reviewed and revised by the experts including a professor of psychiatry and two professors of clinical psychology. The second draft thus produced was back-translated into English by a native English speaker blind to the original version of the WHODAS forms. The first author and experts reviewed convergences and divergences between the original and back-translated scales to produce the third draft in Turkish, which was then pre-tested in a group of patients at a university psychiatry clinic. In light of the observations during pre-testing, the final versions of the WHODAS 2.0 Turkish forms were decided.

Participants

The clinical sample of the study consisted of 35 inpatients or outpatients at the psychiatry clinic at Adnan Menderes University Research Hospital, ranging in age between 18 and 65, and volunteering to participate. Likewise, the control sample of the study consisted of 35 healthy volunteers within the same age range and with no medical and/or psychiatric illnesses impairing functioning. We used G-Power (Faul et al. 2007, Faul et al. 2009) to estimate minimum sample size that would suffice to test our first validity hypothesis with 0.95 power, that is, to detect a large difference (d≥0.8) between the two samples' disability scores, keeping α as well as β (Type I and II error probability) below 0.05 (Cohen 1988).

Primary clinical diagnosis of each clinical subject was enquired via the automation system in the hospital. Besides, each healthy subject enrolled in the study underwent a psychiatric examination to confirm the absence of any psychiatric illnesses. We also recruited a voluntary relative for each subject, being kin, partner or friend living in the same household and having enough information to rate the

subject's functionality on the WHODAS proxy-rated form. We obtained written informed consent of all the subjects and their proxies in compliance with the ethical approval by the Adnan Menderes University Non-Interventional Research Evaluation Committee (No: 56989545/050.04-278).

Clinical Assessment Tools

World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) is an assessment tool developed by the World Health Organization to measure health and disability in the normal population and in clinical practice. This instrument inquires how hard it is for an individual to perform certain key activities common to several cultures under six domains: (1) cognition, (2) mobility, (3) self-care, (4) getting along, (5) life activities, and (6) participation (Uluğ et al. 2001). Each item exploring the difficulty of the individual in carrying out a certain activity within the last 30 days is scored on a 5-point scale of 0 (none), 1 (mild), 2 (moderate), 3 (severe), and 4 (extreme or cannot do). There exist the 36-item full version, the 12-item short version of the instrument, each of which is available in interviewer- proxy- and self-administered forms. There is also a 12+24-item version available only in interviewer-administered form. The proxy form was designed to be rated by a relative if the patient is unable or unfit to self-rate. A general disability score is obtained by administering either the full or the short version of the instrument whereas a set of domain disability scores are obtained only from the full version. Originally, WHODAS 2.0 items were proposed to be coded and scored according to either simple or complex systems (Üstün et al. 2010). Due to confusing variations between these two systems, however, this proposal has not been received enthusiastically by the researchers in the field. Here we followed Andrews et al. (2009) in coding the item ratings between 0 and 4 (as currently suggested by the WHO website as well) and the average item scoring method of the DSM-5 in computing general and domain disability scores (APA 2013). This method entails taking the average of the pertinent item ratings when computing a disability score, thereby yields a score ranging from 0 to 4 regardless of the number of items considered. Obviously, it provides a practical manner of interpreting and comparing the disability scores obtained from various WHODAS 2.0 scales and subscales composed of varying number of items.

Process

We initially asked each participant to self-rate the 36-item form, and his/her relative to rate the 36-item proxy form. Subsequently, 36-item interviewer form was administered to each participant by the principal author and a second rater jointly attending the same interview with the patient. Thus, a total of 9 clinicians pairing with the principal rater to independently rate the clinical subjects on the basis of joint interviews provided data for investigating interrater agreement. Short 12-item forms were not independently administered as they are covered by the 36-item forms.

Statistical Analysis

Comparisons between the two samples were conducted by means of the t-test or x^2 test. We adopted two different approaches to investigate concurrent validity of the WHODAS general disability scores against our external criterion concerning the subjects' membership to the clinical or control sample. First, we employed the t-test and Cohen's d to examine the effect of group membership on a series of general and domain disability scores. Second, we determined optimal cut-off for each general disability score with the aid of ROC (Receiver Operating Characteristic) analysis, and computed psychometric performance (sensitivity and specificity) of each cut-off applied score in differentiating patients from controls. Also, we employed logistic regression analysis to estimate predictive potential of cut-off applied scores by means of Cox-Snell and Nagelkerke R^2 values as well as DORs (Diagnostic Odds Ratios). We calculated intraclass correlation coefficient (ICC) to estimate the level of agreement among the disability scores yielded by different administrative modes, versions, and interviewers. The internal consistency of the scales and subscales of WHODAS 2.0 forms was examined with Cronbach's alpha and corrected item-total correlation coefficients. Entire analyses for the WHODAS validity and reliability study were conducted using the SPSS 17.0 program. Space limitations forbid us presenting here the data and the graphics concerning ROC analyses and item-total correlations in detail, which were presented and interpreted elsewhere (Aslan Kunt 2016).

RESULTS

Socio-demographic and Clinical Characteristics of the Subjects

Table 1 summarizes the socio-demographic characteristics of the clinical subjects and the control subjects, and the results of the statistical tests implying no significant difference between the samples in terms of socio-demographic variables.

The primary psychiatric diagnoses of the clinical subjects as classified into DSM-5 diagnostic groups were as follows: schizophrenia and other psychotic disorders in 5 patients (14.3%), bipolar and related disorders in 6 patients (17.1%), depressive disorders in 11 patients (31.4%), anxiety disorders in 5 patients (14.3%), obsessive-compulsive and related disorders in 2 patients (5.7%), trauma and stressor related disorders in 5 patients (14.3%), somatic symptom and related disorders in 1 patient (2.9%).

Table 1. Socio-demographic characteristics of the control and clinical samples

	Control Sample (n = 35)	Clinical Sample (n = 35)	Statistical Test
	n (%)	n (%)	
Gender			
Female	29 (82.9)	24 (68.6)	χ^2 =1.94, df=1, p=0.16
Male	6 (17.1)	11(31.4)	
Marital Status			
Married	27 (77.1)	17 (48.6)	χ^2 =6.12, df=1, p=0.13
Not married	8 (22.9)	18 (51.4)	
Employment status			
Employed	22 (62.9)	13 (37.1)	χ^2 =4.63, df=1, p=0.31
Unemployed	13 (37.1)	22 (62.9)	
Age (Mean ± sd)	39.54±12.46	36.67±12.31	t=0.98,df=68, p=0.33
Education Year (Mean ± sd)	11.11±5.40	10.40±3.36	t=0.67, df=56.95, p=0.51

Criterion (Concurrent) Validity of Disability Scores

Results of a series of t-test revealed that the subjects' membership in the clinical or control sample has a significant ($p < 0.001$) effect on their general disability scores regardless of the employed WHODAS 2.0 version or administrative mode (Table 2). Computed Cohen's d values suggested that clinical group's mean general disability scores were around two standard deviations higher than those of the control group (Table 3). Likewise, comparisons between the two samples with respect to domain disability scores revealed a large difference in each case as implied by Cohen d's ranging from 0.84 to 3.02 (Table 2 and 3). Overall, these findings support the concurrent validity of the WHODAS general disability as well as domain disability scores against the external criterion.

Table 2 displays sensitivity and specificity of the WHODAS general disability scores at optimal cut-offs revealed by the ROC analysis. Table 3 presents the Cox-Snell and Nagelkerke R^2 values yielded by the logistic regression analyses employed to investigate the value of general disability score status of the subjects (above or below the cutoff) in predicting their group (clinical or control sample) membership. Because Cox-Snell R^2 estimates are usually lower than Nagelkerke R^2 estimates, both were considered in interpreting the size of the prediction, and both suggested a large predictive effect of the general disability scores yielded by any WHODAS version or administration mode ($R^2 = 0.27 – 0.59$).

Likewise, the estimated DORs indicated that subjects with a general disability score above the cut-off point are 12 to 43 times more likely to be a member of the clinical sample (Table 3). Overall, these findings confirm our second validity hypothesis and provide additional evidence for the criterion validity of the entire general disability scores yielded by the interviewer-, self- or proxy-rated forms of the full and short versions of WHODAS 2.0.

Convergent Validity of the Disability Scores

Table 3 displays the ICC coefficients computed separately for each disability score set in each sample to examine the extent of convergence among the scores obtained through three different administration modes. The ICCs concerning the clinical sample were entirely over 0.60 whereas those concerning the control sample were entirely below 0.60 benchmark. These findings provide strong validity evidence for the disability scores of the clinical sample, but mostly questionable evidence for those of the normal sample.

We also investigated convergence between the general disability scores produced by an identical administration mode of the full and short versions. ICC's estimated separately in the control and clinical samples were .84 and .97 for the pairs of scores yielded by the interviewer-administered forms, .95 and .97 by the self-administered forms, and .93 and .97 by the proxy-administered forms, respectively. These validity coefficients suggest strong agreement between the full and short versions of the WHODAS 2.0.

Internal Consistency of the WHODAS Scales

The data summarized in Table 3 suggests high internal consistency of the WHODAS full version general disability scores either in the clinical or in the control sample. Internal consistency of the short versions, however, proved to be mostly sufficient only in the clinical sample. Furthermore, domain disability scales included exclusively in the full versions of the instrument appear to be associated with mostly sufficient levels of internal consistency in the clinical sample, but with mostly questionable or unacceptable levels of consistency in the normal sample. These findings suggest that valid and reliable disability assessment is possible in the clinical population with almost a full range of WHODAS scores, while it is possible in the normal population with only general disability scores yielded by the WHODAS full version.

Inter-rater Agreement

Given the benchmarks for the ICC as a coefficient of validity and/or reliability (Kraemer et al. 2012, Clarke et al. 2013, Shrout and Fleiss 1979), the pertinent data on the Table 3 reveals almost perfect convergence for a full range of WHODAS scores by the two clinicians' performing separate ratings based on a joint interview with each participant in the clinical sample.

Table 2. WHODAS 2.0 disability scores yielded by the interviewer-, self- and proxy-administered forms in Turkish: Comparison between control and clinical samples (t-test), and psychometric performance at optimal cut-offs

Disability Score Varieties	Control Sample		Clinical Sample		Statistics			Cut-off Score	Sensitivity	Specificity
	X	sd	X	sd	t	df	p			
WHODAS-36 General Disability										
36-item – IA Form	1.33	0.26	2.39	0.88	6.85	39.78	<0.001	1.47	0.83	0.77
36 item – SA Form	1.31	0.28	2.47	0.94	6.98	39.93	<0.001	1.63	0.80	0.91
36 item - PA Form	1.29	0.26	2.20	0.81	6.34	40.73	<0.001	1.43	0.89	0.83
WHODAS-12 General Disability										
12-item – IA Form	1.33	0.32	2.28	0.84	6.27	43.67	<0.001	1.43	0.86	0.74
12- item –SA Form	1.32	0.30	2.36	0.93	6.29	41.25	<0.001	1.48	0.80	0.74
12- item - PA Form	1.29	0.28	2.17	0.82	6.02	41.73	<0.001	1.45	0.89	0.77
Cognition Domain										
Cognition – IA Form	1.45	0.41	2.33	0.93	5.13	46.85	<0.001			
Cognition – SA Form	1.28	0.33	2.27	0.94	5.88	42.00	<0.001			
Cognition - PA Form	1.30	0.41	2.02	0.91	4.31	47.51	<0.001			
Mobility Domain										
Mobility - IA Form	1.38	0.52	2.25	1.08	4.26	49.06	<0.001			
Mobility - SA Form	1.40	0.62	2.35	1.19	4.18	51.03	<0.001			
Mobility - PA Form	1.37	0.64	1.98	0.91	3.28	60.66	=0.002			
Self-care Domain										
Self-care - IA Form	1.06	0.12	1.88	0.81	5.90	35.54	<0.001			
Self-care - SA Form	1.11	0.24	1.97	0.93	5.26	38.37	<0.001			
Self-care - PA Form	1.11	0.22	1.64	0.85	3.55	38.56	=0.001			
Getting along Domain										
Getting along - IA Form	1.23	0,33	2.32	0.96	6.30	42.07	<0.001			
Getting along - SA Form	1.28	0.36	2.53	1.18	5.99	40.44	<0.001			
Getting along - PA Form	1.24	0.35	2.10	0.94	5.08	43.36	<0.001			
Life activities Domain										
Life activities - IA Form	1.47	0.52	2.49	1.38	4.09	43.65	<0.001			
Life activities - SA Form	1.40	0.59	2.55	1.25	4.94	48.71	<0.001			
Life activities - PA Form	1.24	0.32	2.30	1.22	4.96	38.59	<0.001			
Participation Domain										
Participation - IA Form	1.30	0.27	2.78	0.87	9.61	40.58	<0.001			
Participation - SA Form	1.33	0.32	2.86	0.92	9.30	41.96	<0.001			
Participation - PA Form	1.39	.40	2.75	0.76	9.36	51.45	<0.001			

Abbreviations
IA: interviewer-administered,
SA: self-administered,
PA: proxy-administered

Table 3. Summary of the collected evidence for validity and reliability of the WHODAS 2.0 disability scores yielded by the interviewer-, self- and proxy-administered forms in Turkish

WHODAS 2.0 Scores	Concurrent Validity *			Convergent Validity		Internal Consistency		Inter-rater Agreement
	Cohen d	R2	DOR	Contrl ICC	Clinic ICC	Contrl α	Clinic α	Clinic ICC
WHODAS-36 General Disability Scores				0.59	0.75			
36-item – IA Form	2.17	0.33 – 0.43	17			0.88	0.97	0.997
36 item – SA Form	2.21	0,44 – 0.59	43			0.90	0.97	
36 item - PA Form	1.99	0.43 – 0.58	37			0.87	0.97	
WHODAS-12 General Disability Scores				0.55	0.70			
12-item – IA Form	1.90	0.33 – 0.43	17			0.77	0.90	0.995
12- item –SA Form	1.96	0.27 – 0.36	12			0.71	0.90	
12-item - PA Form	1.86	0.38 – 0.51	26			0.64	0.90	
Cognition Domain Scores				0.43	0.65			
Cognition – IA Form	1.50					0.65	0.97	0.996
Cognition – SA Form	1.81					0.75	0.89	
Cognition-PA Form	1.25					0.82	0.90	
Mobility Domain Scores				0.55	0.73			
Mobility – IA Form	1.22					0.82	0.88	0.990
Mobility - SA Form	1.17					0.89	0.90	
Mobility- PA Form	0.84					0.88	0.84	
Self-care Domain Scores				0.28	0.76			
Self-care - IA Form	1.98					0.12	0.74	0.990
Self-care - SA Form	1.70					0.15	0.76	
Self-care - PA Form	1.14					0.28	0.83	
Getting along Domain Scores				0.51	0.64			
Getting along - IA Form	1.94					0.71	0.77	0.985
Getting along - SA Form	1.88					0.76	0.88	
Getting along - PA Form	1.54					0.63	0.85	
Life activities Domain Scores				0.59	0.73			
Life activities – IA Form	1.24					0.89	0.97	0.999
Life activities- SA Form	1.42					0.90	0.93	
Life activities- PA Form	1.60					0.85	0.94	
Participation Domain Scores				0.46	0.62			
Participation - IA Form	3.02					0.46	0.82	0.982
Participation - SA Form	2.87					0.67	0.83	
Participation - PA Form	2.61					0.74	0.81	

Abbreviations
Contrl : control sample, Clinic : clinical sample, IA : interviewer-administered , SA : self-administered, PA : proxy-administered, DOR : diagnostic odds ratio, ICC : intraclass correlation coefficient
* The effects sizes revealed by the t-tests are presented in the column titled Cohen d, and those revealed by the logistic regression analyses in the two columns titled R^2 and DOR. For each regression analysis, both Cox-Snell R^2 and Nagelkerke R^2 estimates are displayed respectively.
Color codes for validity/reliability coefficients: red= unacceptable; yellow = questionable/acceptable; green = satisfactory

Table 4. The WHODAS 2.0 domain scores in various studies: Sample means (percentage of maximum score) and internal consistency coefficients (Cronbach alpha)

Study	Country	Form	Sample	CN	MO	SC	GA	LA	PN
Mean sample domain scores (percentage of maximum score)									
AslanKunt&Dereboy (2018)	Turkey	SA	Mixed psychiatric diagnoses (n = 35)	57	59	49	63	64	72
Guilera et al. (2015)	Spain	IA	Bipolar Disorder (n = 291)	56	33	20	79	53	50
Keeley et al. (2014)	USA	SA	Mixed psychiatric diagnoses (n = 99)	37	39	23	35	33	39
Uluğ et al. (2001)	Turkey	IA	Schizophrenia (n = 60)	34	15	14	32	30	33
Eren et al. (2007)	Turkey	IA	Schizophrenia (n = 50)	42	20	22	40	47	43
Akinsulore et al. (2015)	Nigeria	SA	Schizophrenia (n = 100)	25	7	3	45	21	48
Aloba et al. (2015)	Nigeria	SA	Schizophrenia, Bipolar, Depressive Disorders (n = 327)	31	28	28	30	14	31
Internal consistency of the domain scores (Cronbach alpha)									
AslanKunt&Dereboy (2018)	Turkey	SA	Mixed psychiatric diagnoses (n = 35)	0.89	0.90	0.76	0.88	0.93	0.83
Guilera et al. (2015)	Spain	IA	Bipolar Disorder (n = 291)	0.88	0.84	0.73	0.85	0.92	0.90
Silva et al. (2013)	Portugal	IA	Musculoskeletal pain (n = 204)	0.83	0.79	0.84	0.85	0.79	0.80
Carlozzi et al. (2015)	USA	SA	Huntington's disease (n = 477)	0.82	0.89	0.90	0.74	0.83	0.74
Chiu et al. (2014)	Taiwan	IA	Individuals living in elderly or disability institutions (n = 307)	0.84	0.88	0.82	0.88	0.95	0.73

Abbreviations: CN: Cognition Domain, MO: Mobility Domain, SC: Self Care Domain, GA: Getting Along With People Domain, LA: Life Activities Domain, PN: Participation Domain, IA: Interviewer-adminstered, SA: Self-administered

DISCUSSION

Among several WHODAS 2.0 scoring systems, we picked the "average item rating" recommended by the DSM-5 on the basis of the field trial data supporting reliability, convenience and clinical usefulness of the general and domain disability scores obtained with this method (American Psychiatric Association 2013). There are practical advantages of the disability scores representing average item ratings when compared with those representing the sum of item ratings: first, the relative ease of comparisons between the scores that involve varying number of items, second, the ease of interpretation of a variety of disability scores with no need to keep in mind score-specific benchmarks. Nevertheless, the DSM-5 coding proposal that each WHODAS item be rated on a scale ranging from 1 to 5 renders the average item scores in the same range, hence a null score is unavailable even for those with no disability at all. To tackle this problem, the WHODAS items have been rated on a scale ranging from 0 to 4 in various studies (Chi et al. 2014, Andrews et al. 2009, Almazan-Isla et al. 2014, Carlozzi et al. 2015), and in the present study as well.

Our data suggest that the general disability scores yielded by the 36-item WHODAS forms in Turkish are valid and reliable in the clinical and in the control sample. Owing to the time-consuming nature of the interviewer-administered forms, proxy- and/or self-administered forms might be preferable in large studies. The observed convergence among the general disability scores obtained through varying administration modes supports this recommendation to a large extent for the clinical samples (ICC = 0.75), and to some extent for the normal samples (ICC = 0.59). Furthermore, the general disability scores obtained from any 36-item WHODAS form seem to be convenient not only for research purposes but also for clinical purposes, i.e., for individual assessments, given the

internal consistency coefficients exceeding .90 benchmark in the clinical sample of our study (Nunnally and Bernstein, 1988).

Regarding the validity and reliability of the general disability scores yielded by three administration modes of 12-item WHODAS 2.0, we obtained satisfactory or high estimates in the clinical sample, whereas questionable estimates in the normal sample. Though the short interview form scores were not as reliable as the long interview form scores, they were still high enough for either research purposes or clinical evaluations. Furthermore, the performance of the short interview form scores in discriminating the clinical subjects from normal subjects was no different than that of the full version interview scores. The short versions of the self- or proxy-rated forms, however, were comparably less efficient than their full version counterparts in producing disability scores which are discriminative and reliable in both samples.

A notable finding of our study is the astounding variation of the psychometric properties of the domain disability scores between the clinical sample and the normal sample. That is, the validity coefficients estimated for each of the six domain scores were satisfactory in the clinical sample but not in the control sample (Table 3). Likewise, the alpha coefficients we computed in the normal sample fell well below the acceptable limits. These findings are consistent with the findings of Guilera et al. (2015), which imply some difficulty in detecting the slight differences among disability degrees of the subjects whose functioning is close to adequate level.

Table 4 displays mean WHODAS domain scores expressed as the percentage of maximum possible score and pertinent Cronbach's alpha coefficients observed in various studies from different countries. In our study, the degree of disability in the clinical sample was comparably less enhanced in the self-care and mobility domains, and more enhanced in the life activities and getting along domains. These findings replicate results of the original WHODAS 2.0 study coordinated by the WHO (Üstün et al. 2010), and are in line with the previous research reporting relatively less severe disability in the self-care domain among schizophrenic patients (Uluğ et al. 2001, Eren et al. 2007, Akinsulore et al. 2015). Likewise, mostly satisfactory internal consistency estimates concerning domain scores yielded by the self-administered forms in our study are in accordance with the previous reports summarized in Table 4. The inter-rater reliability of the WHODAS 2.0 scores was estimated as nearly perfect in our clinical sample, which is partly due to that the interviewer-administered forms were rated by two clinicians simultaneously during or immediately after a joint interview with each clinical participant.

Owing to the absence of cut-off points for the detection of significant disability across cultures, a variety of cut-offs have been applied in different studies. Andrews et al. (2009) reported that a clinically significant degree of disability was associated with the WHODAS 2.0 short form scores of 10 or higher as computed with the simple scoring system. Another study investigating risk factors associated with disability in diabetic patients (Von Korff et al. 2005) considered 12-item WHODAS-II scores over 45 percent of the maximum possible score as indicating significant disability. Our data suggest that an average item rating disability score roughly equating to or exceeding a score of 1.5 imply significant loss of functionality associated with psychopathology when using the full or short version of WHODAS 2.0 in mental health care facilities in Turkey. Researchers and clinicians might refer to Table 2 for the exact cut-offs that could be applied to the general disability scores obtained through different WHODAS forms.

Limitations and strengths of the study

The major limitation of this study stems from the fact that each of the clinical and control samples consisted of 35 subjects. Thus, relatively small number of participants recruited at a single center might have constrained our samples' capacity to represent population characteristics adequately, which in turn restricted the generalizability of our findings. Nonetheless the sample size of this study was determined to test, with type I and type II error rates of 5% at most, our one-tailed validity hypothesis that the WHODAS disability scores of the clinical sample were to be higher than those of the control sample, and the difference was to be large ($d \geq .80$) at each comparison. Accordingly, the number of participants in either sample was enough for between group comparisons employed to test this hypothesis.

The strength of the present study is related to its capacity to provide psychometric data for all types of disability scores which could be obtained by administering the WHODAS 2.0 forms in Turkish. Hence, we were able to examine in a comparative manner the validity and reliability estimates pertaining to the general disability and domain disability scores yielded by three different administration modes of the full and short versions of the instrument.

CONCLUSIONS

The Turkish translation of the WHODAS 2.0 full version, whether interviewer-, proxy- or self-administered, yield general disability scores which could be used not only for research purposes involving assessment of clinical and/or normal samples, but also for clinical purposes involving individual evaluations. Accordingly, researchers might prefer to use self- or proxy-administered forms instead of the interviewer-administered forms to assess general disability in large samples with less time and effort. On the other hand, the short WHODAS 2.0 forms yield general disability scores

of which validity and reliability being estimated as sufficient in the clinical sample, whereas as questionable/acceptable in the normal sample. Therefore, we recommend that the researchers use long forms to assess general disability in normal samples. Likewise, clinicians aiming to assess general disability in individual patients by means of the self- or proxy-administered forms might prefer to use full versions of these forms that yield scores with better discriminative properties than those yielded by the short versions. If the individual assessment is to be performed via interviewer-administered forms, clinicians might prefer to use the short version yielding equally discriminative general disability scores with less time and effort as compared with the full version.

As to the disability assessment in various domains in the Turkish clinical population, our data suggest that the entire WHODAS 2.0 domain scores can legitimately be used for research purposes; yet the reliability of some domain scores may not be high enough for use with clinical purposes. We recommend that clinicians consider this constraint when using the domain scores for clinical evaluation of individual patients. Furthermore, we warn against the use of the domain scores in the normal population on the basis of the validity and reliability estimates of some domain scores that dropped far below acceptable levels in our control sample.

REFERENCES

Akinsulore A, Mapayi BM, Aloba OO et al (2015) Disability assessment as an outcome measure: a comparative study of Nigerian outpatients with schizophrenia and healthy control. Ann Gen Psychiatry 14:40.

Almazan-Isla J, Comin-Comin M, Damian J et al (2014) Analysis of disability using WHODAS 2.0 among the middle-aged and elderly in Cinco Villas, Spain.Disabil Health J 7:78-87.

Aloba O, Akinsulore A, Mapayi B et al (2015) The Yoruba version of the Beck Hopelessness Scale: psychometric characteristics and correlates of hopelessness in a sample of Nigerian psychiatric outpatients. Compr Psychiatry 56:258-71.

American Psychiatric Association (2013) The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. American Psychiatric Association; Arlington, VA.

American Psychiatric Association (2000) The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. American Psychiatric Association.

Andrews G, Kemp A, Sunderland M et al (2009) Normative data for the 12 item WHO Disability Assessment Schedule 2.0. PloS one 4: e8343.

AslanKunt D (2016) Validity and reliability of the World Health Organization Disability Assessment Schedule (WHODAS 2.0) in Turkish. Dissertation, Adnan Menderes University, Aydın.

Carlozzi NE, Kratz AL, Downing NR et al (2015) Validity of the 12-item World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) in individuals with Huntington disease (HD).Qual Life Res 24:1963-71.

Chen H, Cohen P, Chen S (2010) How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies.Communications in Statistics—Simulation and Computation, 39: 860-4.

Chi WC, Chang KH, Escorpizo R et al (2014) Measuring disability and its predicting factors in a large database in Taiwan using the World Health Organization Disability Assessment Schedule 2.0. Int. J Environ Res Public Health 11:12148-61.

Chiu TY, Yen CF, Chou CH et al (2014) Development of traditional Chinese version of World Health Organization disability assessment schedule 2.0 36-item (WHODAS 2.0) in Taiwan: validity and reliability analyses. Res DevDisabil 35:2812-20.

Clarke DE, Narrow WE, Regier DA et al (2013) DSM-5 field trials in the United States and Canada, Part I: study design, sampling strategy, implementation, and analytic approaches. Am J Psychiatry 170:43-58.

Cohen J (1988) Statistical Power Analysis for the Behavioral Sciences. (2nd Ed). Hillsdale, New Jersey: Lawrence Erlbaum.

Druss BG, Marcus SC, Rosenheck RA et al (2000) Understanding disability in mental and general health conditions.Am J Psychiatry 157:1485-91.

Eren I, Simsek D,Caliskan AM (2007) Effects of disability and symptom severity on quality of life in schizophrenia patients. Düşünen Adam 20:68-78.

Faul F, Erdfelder E, Buchner A et al (2007) G Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175-91.

Faul F, Erdfelder E, Buchner A et al (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav Res Methods 41:1149-60.

Glas AS, Lijmer JG, Prins MH et al (2003) The diagnostic odds ratio: a single indicator of test performance. J ClinEpidemiol 56:1129– 35.

Guilera G, Gómez-Benito J, Pino Ó et al (2015) Disability in bipolar I disorder: The 36-itemWorld Health Organization Disability Assessment Schedule 2.0. J Affect Disord 174:353-60.

Keeley JW, Flanagan EH, McCluskey DL (2014) Functional impairment and the DSM-5 dimensional system for personality disorder. J PersDisord 28:657-74.

Kraemer HC, Kupfer DJ, Clarke DE et al (2012) DSM-5: How reliable is reliable enough? Am J Psychiatry 169:13-50.

Nunnally JC, Bernstein IH (1994) Psychometric Theory. McGraw Hill Series in Psychology (3rd. Ed), New York.

Sheehan DV, Harnett-Sheehan K, Raj BA (1996) The measurement of disability. IntClinPsychopharmacol11:89-95.

Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420-8.

Silva C, Coleta I, Silva AG et al (2013) Adaptation and validation of WHODAS 2.0 in patients with musculoskeletal pain. Rev Saude Publica 47:752-8.

Sireci SG, Sukin T (2013) Test validity. In: APA Handbook of Testing and Assessment in Psychology (Vol. 1), KF Geisinger, BA Bracken, JF Carlson, JIC Hansen, NR Kuncel, SP Reise, MC Rodriguez (Eds). Washington, DC: American Psychological Association: 61-84

Uluğ B, Ertuğrul A, Göğüş A et al (2001) Reliability and validity of the Turkish version of the World Health Organization Disability Assessment Schedule-II (WHO-DAS-II) in schizophrenia. Turk Psikiyatri Derg 12:121-30.

Üstün TB, Chatterji S, Kostanjsek N et al (2010) Developing the World Health Organization Disability Assessment Schedule 2.0. Bull World Health Organ 88: 815–23.

Üstün TB, Kostanjsek N, Chatterji S et al (2010) Measuring Health and Disability: Manual for WHO Disability Assessment Schedule (WHODAS 2.0). Accessed on Aug 03, 2015 athttp://www.who.int/classifications/icf/whodasii/en/

Von Korff M, Katon W, Lin EH et al (2005) Potentially modifiable factors associated with disability among people with diabetes. Psychosomatic Medicine 67: 233-40.

World Health Organization.Process of translation and adaptation of instruments. Accessed on Aug 07, 2015 athttp://www.who.int/substance_abuse/research_tools/translation/en/

World Health Organization.WHO Disability Assessment Schedule 2.0 (WHODAS 2.0). Accessed on Aug 07, 2015 at http://www.who.int/classifications/icf/whodasii/en/