





Research

Psychometric properties of full WHODAS 2.0 questionnaire for application through phone surveys: an experience in PERSIAN traffic cohort

Nasrin Shahedifar^{1,2}  · Homayoun Sadeghi-Bazargani¹  · Mina Golestani¹  ·
Mohammad Asghari Jaafarabadi^{1,3,6}  · Faramarz Poursghar⁴  · Shahrzad Bazargan-Hejazi⁵ 

Received: 25 August 2024 / Accepted: 27 March 2025

Published online: 19 April 2025

© The Author(s) 2025 [OPEN](#)

Abstract

Introduction To promote trauma survivors' care quality, assessment of their disability is essential. This study assessed psychometric characteristics of World Health Organization disability assessment schedule (WHODAS 2.0) and its applicability in screening via phone survey through advanced methods.

Methods We recruited 321 adult patients within a national PERSIAN Traffic Cohort: post-crash phase. They completed 36-item WHODAS 2.0. The translation validity– with WHO permission– was examined. The confirmatory factor analysis (CFA) tested the dimensionality, reliability and validity using smart partial least squares structural equation modeling. Item properties – discrimination, difficulty, and differential item functioning – were assessed with the item response theory-graded response model. Test-retest reliability was evaluated by retesting 40 participants after 9 days.

Results The majority of 321 respondents were men (72%) or employed (75%). The mean age was 40.3(SD=13.4;19- 70yrs.). The translation validity was supported. CFA confirmed convergent validity, discriminant validity, predictive accuracy, predictive relevance, and predictive quality. The composite reliability, Cronbach's alpha (0.94-0.95) and rho-A, intra class correlation coefficient, and weighted Kappa coefficient supported reliability. The 6-factor model showed sufficient fit (GoF=0.58). Discrimination parameters distinguished respondents by disability levels, except for three items. The test characteristic and information function curves illustrated items' difficulty shifting toward higher disability level. Two items specified significant differential functioning by sex.

Discussion The findings support reliability and validity of the Persian WHODAS 2.0. The findings demonstrate that the measure is especially apt at distinguishing higher-levels of disability. This highlights the tool's value in generating functional profiles to assess disability among trauma survivors in both clinical and research settings.

Homayoun Sadeghi-Bazargani and Mina Golestani have been contributed equally to this work.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1186/s12982-025-00536-3>.

✉ Homayoun Sadeghi-Bazargani, homayoun.bazargani@gmail.com; ✉ Mina Golestani, mgolestani8958@gmail.com; ; Nasrin Shahedifar, swshahedi@gmail.com; Mohammad Asghari Jaafarabadi, m.asghari862@gmail.com; Faramarz Poursghar, poursgharf@tbzmed.ac.ir; Shahrzad Bazargan-Hejazi, shahrzad2736@gmail.com | ¹Road Traffic Injury Research Center, Tabriz University of Medical Sciences, 5167846311, Tabriz, East Azerbaijan, Iran. ²Injury Epidemiology and Prevention Research Group, Turku Brain Injury Center, Turku University Hospital and University of Turku, Turku, Finland. ³Cabrini Research, Cabrini Health, VIC 3144, Australia. ⁴Department of Medical Informatics, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran. ⁵College of Medicine, Charles Drew University of Medicine and Science, Los Angeles, CA, USA. ⁶School of Public Health and Preventative Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, VIC 3800, Australia.



Keywords Disability · Epidemiologic methods · Traffic accidents · Accidents · Factor analysis · Disability evaluation · Statistical models

1 Introduction

Globally, 16% of total population experiences a major disability. Severely disabled people die up to 20 years earlier than those without such conditions [1]. This imposes financial burden on society and family [2]. Trauma injuries are one of the main causes of disability in active population, particularly in low- and middle- income countries [3], residence for 80% of world's disabled population [2]. The majority of traumatized people (80%) suffer from mild disability [4]. In Iran, trauma caused by road accidents was the second top cause of disability and the third principal cause of death [5], amounting to a burden equals to 12,478 years, 20% of which is due to disability [6]. It is crucial to assess neglected post-trauma disabilities early and precisely to provide timely health [7] and rehabilitation services.

To reach a concise estimation of disability, a complete tool is required. Among various tools, World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0), available in 36- and 12-item versions [8], is well designed for diverse conditions [9]. It is integrated into WHO resolution for 11 th International Classification of Disease–ICD- 11 to develop functioning profiles and disability in clinical assessment [10–12]. Its psychometric properties have been broadly investigated across different diseases and methodologies [13]. However, only the short version has been assessed in post-crash trauma survivors [14]. None of existing Persian translations have been validated via phone surveys among high-risk trauma survivors with road traffic accidents [15–18]. According to the above-mentioned statements, our expert team selected WHODAS 2.0 to apply within the PERSIAN cohort study setting [19].

The paucity of the full version's comprehensive psychometric evaluations by factor analysis and item-response theory, particularly in among traumatized population, along with the adverse role of road traffic accidents (RTA) in causing severe disability in Iran [20] led us to conceptualize and manage to deeply assess its psychometric characteristics.

2 Methods

2.1 Study participants

We assessed the psychometric properties applying phone interviews in clinical population. The study population comprised trauma survivors who hospitalized after RTA in one of two referral trauma centers and enrolled within the post-crash PERSIAN traffic safety and health cohort platform. Eligible survivors were aged ≥ 18 yrs, and classified as high-risk patients based on PTSD and PHQ Scores. A representative sample of them were invited to complete WHODAS 2.0. The participation and data collection are described more elsewhere [21].

The inclusion criteria were: being hospitalized after RTA, registered, aged ≥ 18 yrs, high-risk participants, and providing their informed consent. Those unable to complete the questionnaire due to any disorders were excluded. Individuals with cognitive impairments were not systematically excluded from the study unless they were unable to complete the study assessments. Our questionnaires' completeness was guaranteed since an experienced interviewer collected data [22].

2.2 Study procedure

We took several steps to assess psychometric properties of the WHODAS 2.0 as detailed below:

2.2.1 Translation

In order to develop the Persian version of WHODAS 2.0, the WHO permission was requested and issued for the second time and submitted to the Road Traffic Injury Research Center (permission ID:380,480). An expert team conducted translation based on WHO guidelines: (1) forward translation; (2) Expert panel; (3) Back-translation; (4) Pre-testing and cognitive interviewing, and (5) Final version [14].

The forward translation to Persian language was done by two qualified translators as native-like English speakers familiar with health and disability terms. The back-translation process by a linguist, blind to the original phrases, was completed. Additionally, a semantic challenge, the complexity of item 6.5 ("emotionally affected") was faced. The expert team resolved it after discussions with translators, and implementing minor revisions. Then, the translated questionnaire was administered to several individuals as a pre-test to ensure the final questionnaire's compatibility and comprehensibility.

2.2.2 Reliability

The scale score reliability was evaluated by test–retest method, with a 9-day average interval between measurements (range: 8–10 days), least which reduces learning effects [23, 24]. The participants had no changes in their drugs and treatment during that time interval. We took consistency, fixed protocol for testing, and one tester, then, the sources of inconsistencies were managed.

2.2.3 Validity

We applied confirmatory factor analysis (CFA) to test the dimensionality of the original construct. Then, item response theory (IRT) models study how each individual item or the set of items relates to the trait of interest [25] and to appraise the performance of the scale in a context [26].

2.2.4 Sample size and recruitment

The minimum sample size of 29 to measure intra-class correlation coefficient (ICC) is estimated using R package of "calculate ICC sample size" (Calculate ICC Sample Size ($p_0 = 0.6$, $k = 2$, $\alpha = 0.05$, $\text{tails} = 2$, $\text{power} = 0.95$)) [27]. So, we called 40 patients in both gender, four age groups (18–29; 30–44; 45–64; 65 and above), and four educational levels (illiterate, up to six classes, 7–12 classes, associate degree and higher). Some patients rejected to participate (five), died (one), and did not answer the call (two). They were substituted by the same sex, age group, and education level.

To assess the validity of the instrument, based on "a-priori sample size calculator for structural equation models", with $F^2 = 0.33$ [28], $\text{power} = 0.80$, and $\alpha = 0.05$ for 36 observed variables, a sample size of 200 was recommended [29]. Then, we took 321 patients (about 50 cases per factor in CFA), and the model fit would maximize.

2.3 Measures

The questionnaires were demographic information and WHODAS 2.0

2.3.1 World health organization disability assessment schedule

As a general measure, WHODAS 2.0 was developed based on the framework of International Classification of Functioning, Disability and Health to measure disability during the past 30 days [30]. It includes six domains: cognition (D1;6 items), mobility (D2;5 items), self-care (D3;4 items), getting along (D4;5 items), life activities (D5;8 items), and participation (D6;8 items) [16, 31].

The response choices ranged from zero (none) to four (extreme/cannot do). The scale score ranged from 0 (no disability) to 144 (complete disability). Higher the score, higher the disability level is [32]. We transformed the raw scores into a range from 0 to 100 normalized using the min–max normalization technique via the following formulae:

$$\text{Min} - \text{max Normalized score} = (\text{raw score} - \text{minimum}) \div (\text{possible range of 144}) \times 100$$

The subscale scores were calculated by summing the items in each subscale [33]. Disability severity was categorized as no problem (0–4), mild (5–24), moderate (25–49), severe/extreme (50–100) [34]. Surveys with two or more missing items was excluded, while for one missing item, the mean of the remaining items was assigned [9].

2.4 Statistical analysis

The descriptive analysis and IRT analysis were performed in STATA statistical package version 15 [35]. The CFA were performed in structural equation model of Smart Partial Least Squares (PLS) 3.3.9 program [36], because the current model is quite simple, with a limited number of observed variables [37]. P-values less than 0.05 were considered as significant.

2.4.1 Reliability

Internal consistency reliability was assessed by Cronbach's Alpha [38] as moderate (0.5–0.75), good (0.75–0.9), and excellent > 0.90 [26, 39]. Composite reliability (CR) also assessed the internal consistency of the latent construct of disability [40]. The stability over time was measured by ICC at full scale and subscale levels [41]. ICC was estimated via the "single-rater type ($k = 1$), absolute-agreement, 2-way mixed-effects model with 95% confidence intervals [39]. Poor-to-fair ICCs are ≤ 0.4 , moderate ICCs: 0.41–0.60, good ICCs: 0.61–0.80 and excellent ICCs > 0.80 [42]. The reliability was estimated by standard error of measurement (SEM). The smallest detectable change (SDC) was also measured using standard error of measurements ($SDC = 1.96 \times \sqrt{2} \times SEM$) [43]. The weighted kappa is applied for each ordinal item. A recommended kappa of ≥ 0.60 specifies good test–retest reliability [44]. The normality of data to assess the reliability was tested based on values of skewness (± 3) and kurtosis (± 7) [45, 46].

2.4.2 Validity

We assessed model validity using confirmatory factor analysis (CFA), in the Smart PLS algorithm. The normality at item- and total-score levels was assessed using skewness (> 3) and Kurtosis (> 10) [47, 48]. The "factor weighting scheme" was set with Maximum Iterations and Stop Criterion at 1000 and 7, respectively. The outputs are Outer Loadings, R square, Cronbach's Alpha, rho_A (Spearman's rank-correlation coefficient), CR, Average Variance Extracted (AVE), Model fit. Loadings ≥ 0.5 demonstrate adequate item reliability [40]. In reflective measurement models, Cronbach's alpha and CR represent the lower and upper bounds of internal consistency, with rho_A falling between them [49]. An AVE > 0.5 [12] and CR > 0.7 confirm that items measuring similar constraints are loaded to one factor [50]. The HTMT (heterotrait-monotrait ratio) verifies discriminant validity (correlations > 0.90 means no discriminant validity) [51].

The inner model metrics include R square, path coefficients size, F^2 , Q^2 , and significance [52]. R^2 values, indicating predictive accuracy, are 0.26 (substantial), 0.13 (moderate), 0.02 (weak) for endogenous latent variables [53]. Stone-Geisser's Q^2 values, indicating predictive relevance, are 0.02 (small), 0.15 (medium), and 0.35 (large) [54]. PLS Path Modeling lacks an overall fit index, but a global goodness of fit (GoF) criterion, introduced in 2004, evaluates both measurement and structural model performance. The GoF index is obtained as the geometric mean of the average communality index and the average R square values [12] with thresholds of 0.10 (low), 0.25 (moderate), and 0.36 (high) [55].

We applied the graded response model (GRM) for ordered responses under categorical IRT models [25] to estimate items' discrimination and difficulty indices. The discrimination index describes the test sensitivity to differentiate the characteristics severity. The difficulty index identifies the level of a perceived problem needed to achieve a 50% probability of choosing a particular score [56]. The items with discrimination < 0.700 offer little information and not directly relate to the latent trait [57]. The items with $\alpha > 1.0$ are highly discriminant.

Table 1 Reliability indicators of WHODAS 2.0 (n = 40)

Subscales	Mean (SD)		Cronbach's alpha	Composite reliability	ICC* (95% CI)	SEM** (95% CI)	SDC***	Agreement
	Test	Retest						
Cognition	5.31 (11.00)	3.44 (7.89)	0.79	0.93	0.93 (0.86–0.96)	2.53	7.01	Excellent
Mobility	24.13 (29.22)	22.25 (27.80)	0.73	0.97	0.99 (0.98–0.99)	2.83	7.84	Excellent
Self-care	16.09 (19.40)	13.28 (17.46)	0.67	0.93	0.98 (0.93–0.99)	2.60	7.21	Excellent
Getting along	5.88 (9.26)	5.25 (8.69)	0.78	0.86	0.97 (0.95–0.98)	1.55	4.29	Excellent
Life activities	30.23 (25.99)	29.38 (26.36)	0.67	0.99	0.99 (0.99–0.99)	3.68	10.20	Excellent
Participation	32.03 (21.75)	29.53 (20.34)	0.68	0.92	0.98 (0.96–0.99)	2.96	8.20	Excellent
Total score	20.68 (14.67)	18.96 (13.99)	0.76		0.99 (0.94–0.99)	1.43	3.96	Excellent

*ICC Intra-Class Correlation Coefficient; at significance level of 0.0001. **SEM Standard Error of Measurement. ***SDC Smallest Detectable Change

Table 2 Baseline characteristics of study population (n = 321)

Variables	Categories	N = 321 (%)
Sex	Male	233 (72.6%)
	Female	88 (27.4%)
Age	18 - 24 yrs	42 (13.1%)
	25 - 44 yrs	163 (50.8%)
	45 - 64 yrs	99 (30.8%)
	65 yrs and above	17 (5.3%)
Employment	Employed	237 (75.5%)
	Unemployed	77 (24.5%)

Differential item functioning (DIF) evaluates if an item behaves differently across respondents with similar disability [58] and indicates that the probability of endorsement of one response category among similarly disabled patients is different for subgroups of people such as gender, age [58]. It may result from shifts in the discrimination parameter (a-DIF), the difficulty parameter (b-DIF), or both (ab-DIF). To test for uniform (b-DIF) and non-uniform (ab-DIF) DIF [25], items with $P > 0.05$ have no difference between groups. The value above 0.1 confirmed no uniform DIF [59].

3 Results

3.1 Translation validity

In item 6.5, “emotionally affected” with semantic complexity was discussed. The forward and back translations were satisfactorily compatible, as reported by three experts separately on a four-level Likert scale. Modified Kappa rate (> 0.76) indicated the complete compatibility, except for item 6.5. After minor revisions, to see items comprehensibility, the final translation was completed by several respondents (Online Resource 1).

3.2 Reliability

The skewness and kurtosis values (0.46 & 2.77, respectively) specify the total score normality. The total Cronbach’s α coefficient was 0.76 ranged from 0.67 (Life activities) to 0.79 (Cognition), demonstrating moderate to good internal consistency. The CR showed the internal consistency of the latent construct of disability (Table 1).

The ICC examined consistency over time by the single-rater type, absolute-agreement, two-way mixed-effects model with 95% CI, representing excellent reliability at 0.99 (overall), varied between 0.93 (Cognition) and 0.99 (Mobility, Life activities). Item-level ICC varied from 0.85 (good) to 0.97 (excellent). The lowest and highest mean test scores were in Cognition and Mobility, respectively (Table 1). The item-level reliability, measured by weighted Kappa coefficient ranged from 0.66 (D1.1) to 1.00 (D2.4, D3.4) (Online Resource 2). The smallest clinically important difference estimates were 1.43 (estimated as SEM). The minimal detectable change was 3.96 points exceeding the smallest clinically important difference.

3.3 Validity

For 321 respondents, the mean age (SD) was 40.3 (13.4; 19 - 70 yrs). They were mostly male (72%) or employed (75%) (Table 2).

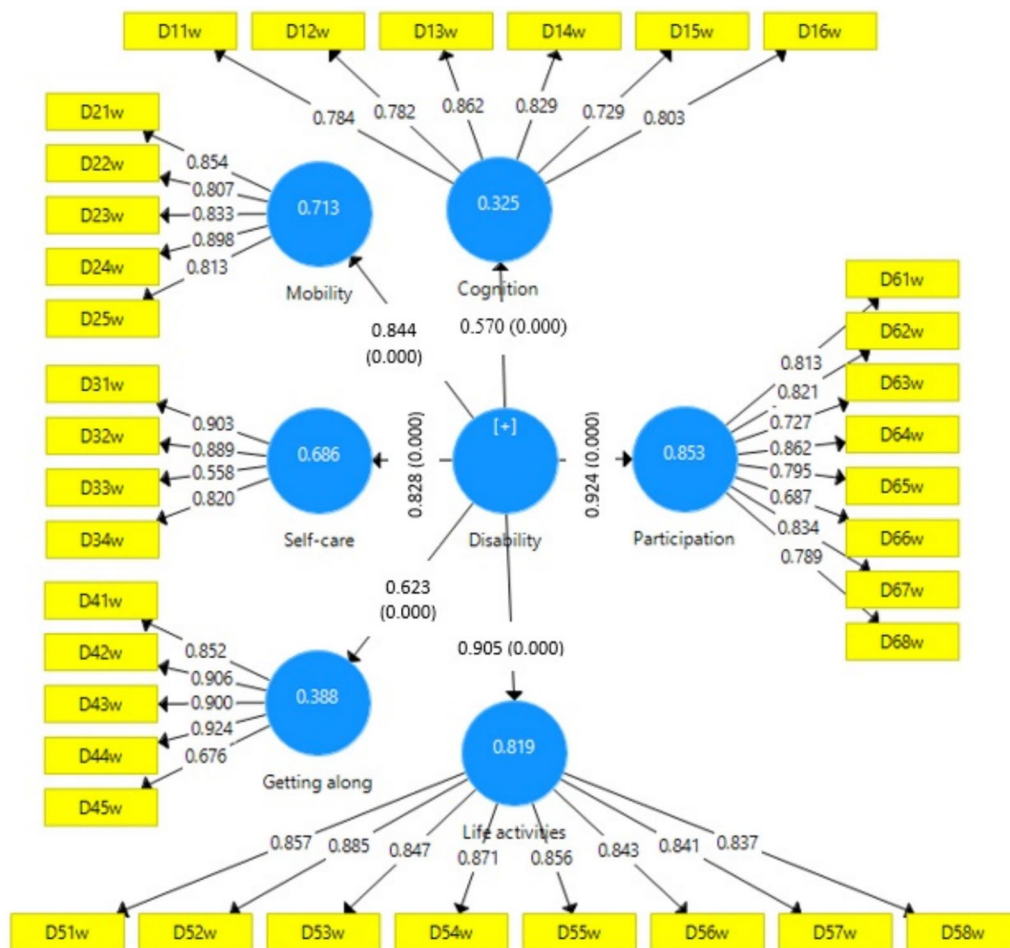


Fig. 1 Output of the WHODAS 2.0 construct in full model test

Table 3 Matrix of latent variables' reliability and validity (n = 321)

Measurement Model							
Scales	D1	D2	D3	D4	D5	D6	
D1- Cognition	0.799	0.441	0.452	0.611	0.408	0.569	
D2- Mobility	0.395	0.842	0.838	0.385	0.791	0.782	
D3- Self-care	0.368	0.729	0.804	0.464	0.782	0.829	
D4- Getting along	0.549	0.349	0.390	0.856	0.444	0.596	
D5- Life activities	0.380	0.737	0.712	0.418	0.855	0.827	
D6- Participation	0.516	0.720	0.724	0.546	0.784	0.793	
Cronbach's alpha	0.886	0.897	0.807	0.905	0.948	0.915	
rho_A	0.893	0.900	0.847	0.909	0.951	0.921	
Composite Reliability	0.914	0.924	0.877	0.932	0.956	0.931	
Average Variance Extracted	0.639	0.708	0.647	0.733	0.731	0.628	
F square	0.481	2.484	2.183	0.633	4.540	5.813	
Cross-validated Commuality	0.493	0.558	0.432	0.598	0.660	0.522	
Cross-validated Redundancy	0.204	0.500	0.433	0.275	0.587	0.528	
Structural Model							
F square	0.481	2.484	2.183	0.633	4.540	5.813	
Cross-validated Commuality	0.493	0.558	0.432	0.598	0.660	0.522	
Cross-validated Redundancy	0.204	0.500	0.433	0.275	0.587	0.528	

3.3.1 Confirmatory factor analysis

Assumptions testing indicated no variables with zero standard deviations. All variables were normally distributed at item and total levels using skewness and Kurtosis.

3.3.1.1 Measurement model Bootstrapping method confirmed convergent validity with significant item factor loadings ($p < 0.0001$). Item loadings ranged from 0.558 (3.3, t-statistics:9.990) to 0.924 (4.4, t-statistics:86.818). Items loadings above 0.5 confirmed item validity. All t-values are above 2.58 and significant ($p < 0.01$) (Fig. 1).

Path model (reflective). The values inside the circles represent the coefficient of determination (R^2). The values overlapping the arrows pointing towards the circles represent the path coefficients (P-values). The values overlapping the arrows pointing towards the yellow rectangular shapes represent the loading values.

The Fornell-larcker criterion reports the discriminant validity. The elements in the matrix diagonal, signifying the AVEs' square roots, are greater than the off-diagonal values in their corresponding row and column. The values above the diagonal, HTMT (< 0.85), confirm discriminant validity. The correlation coefficients, below the diagonal, are significant ($p < 0.01$). AVEs (> 0.5) support the convergent validity (Table 3).

3.3.1.2 Structural model All path coefficients are statistically significant (Fig. 1). The model explains a substantial variance in disability constructs, between 32% (Cognition) to 85% (Participation). The R^2 values indicate model's significant GoF and good predictability.

The F^2 variables ≥ 0.35 are significant. The positive Q^2 values confirm the model's predictive relevance. Construct cross-validated communalities above 0.35 confirm large predictive relevance. Cross-validated Redundancy values for Cognition and Getting along are at medium level, while others are at strong levels (Table 3).

3.3.1.3 Model fit The full model shows GoF of 0.58 (> 0.36) and the Standardized root mean square residual (SRMR) at 0.08 ($P = 0.047$) verifying the quality of the model. Besides, other model fit indices include comparative fit index (CFI) of 0.96 (≥ 0.95), Tucker Lewis index (TLI) of 0.95 (≥ 0.95), root mean square error of approximation (RMSEA) of 0.12 (≤ 0.08), standardized root mean square residual (SRMR) of 0.08 (≤ 0.08), RMS theta of 0.18 ($0.12 < \text{Borderline acceptable fit} < 0.20$), and Chi-square: 3816.43 (non-significant) [60] [61].

Table 4 IRT calibration results of items in full WHODAS 2.0 (n = 321)

Items	Mean (SD)	Item-test correlation	Item-rest correlation	α	β_1 (≥ 1)	β_2 (≥ 2)	β_3 (≥ 3)	β_4 (≥ 4)
D1.1	1.28 (1.34)	0.57	0.54	1.05	-0.26	0.35	1.44	2.85
D1.2	1.25 (1.35)	0.49	0.46	0.85	-0.29	0.51	1.62	3.30
D1.3	1.04 (1.31)	0.57	0.55	1.13	0.19	0.71	1.62	2.90
D1.4	0.95 (1.32)	0.52	0.49	1.02	0.42	0.99	1.66	3.06
D1.5	0.52 (1.02)	0.42	0.40	0.93	1.39	1.96	2.95	4.46
D1.6	0.71 (1.12)	0.51	0.48	1.04	0.74	1.44	2.31	3.92
D2.1	1.85 (1.63)	0.71	0.69	2.16	-0.43	-0.14	0.33	0.89
D2.2	1.67 (1.40)	0.65	0.63	1.73	-0.73	-0.03	0.65	1.64
D2.3	0.82 (1.20)	0.68	0.66	2.63	0.40	0.80	1.38	1.91
D2.4	1.39 (1.53)	0.76	0.74	2.81	-0.08	0.31	0.69	1.21
D2.5	2.10 (1.64)	0.67	0.64	2.05	-0.68	-0.26	0.13	0.63
D3.1	1.22 (1.43)	0.71	0.68	2.57	0.03	0.39	0.90	1.47
D3.2	1.16 (1.34)	0.65	0.62	2.06	-0.01	0.42	1.07	1.90
D3.3	0.68 (1.12)	0.46	0.44	1.10	0.86	1.40	2.39	3.65
D3.4	1.40 (1.65)	0.74	0.72	2.68	0.13	0.30	0.65	1.01
D4.1	0.77 (1.19)	0.50	0.47	0.91	0.71	1.32	2.72	3.41
D4.2	0.75 (1.20)	0.56	0.53	1.12	0.76	1.16	2.07	3.24
D4.3	0.78 (1.13)	0.57	0.55	1.15	0.44	1.13	2.24	3.44
D4.4	0.80 (1.31)	0.58	0.56	1.30	0.73	1.13	1.65	2.41
D4.5	0.92 (1.37)	0.55	0.52	1.18	0.58	0.96	1.72	2.31
D5.1	1.25 (1.50)	0.82	0.80	5.30	0.10	0.38	0.71	1.22
D5.2	1.60 (1.64)	0.83	0.82	5.16	-0.13	0.14	0.48	0.83
D5.3	1.40 (1.62)	0.80	0.78	4.37	0.04	0.27	0.58	0.98
D5.4	1.73 (1.66)	0.81	0.79	4.28	-0.25	0.04	0.37	0.77
D5.5	1.71 (1.61)	0.81	0.79	3.31	-0.42	-0.13	0.29	0.85
D5.6	1.83 (1.61)	0.79	0.77	2.79	-0.56	-0.18	0.19	0.82
D5.7	1.72 (1.65)	0.79	0.77	3.04	-0.37	-0.06	0.28	0.79
D5.8	1.93 (1.63)	0.78	0.76	2.87	-0.63	-0.27	0.11	0.67
D6.1	1.53 (1.52)	0.80	0.78	2.89	-0.20	0.10	0.70	1.11
D6.2	1.40 (1.54)	0.81	0.79	3.38	0.02	0.22	0.64	1.22
D6.3	1.17 (1.39)	0.68	0.66	1.68	-0.01	0.48	1.16	1.82
D6.4	2.01 (1.63)	0.80	0.79	2.88	-0.43	-0.24	0.16	0.79
D6.5	2.16 (1.51)	0.71	0.69	1.66	-1.01	-0.59	0.21	0.92
D6.6	2.36 (1.60)	0.57	0.54	1.36	-1.06	-0.82	-0.22	0.64
D6.7	2.00 (1.61)	0.75	0.72	2.20	-0.58	-0.29	0.15	0.89
D6.8	1.55 (1.54)	0.72	0.69	1.89	-0.24	0.09	0.61	1.36

*P-value < 0.0001; α : Discrimination; The bold items indicate difficulty level shifted towards the elevated disability level compared to the average disability level

3.3.2 Graded response model (GRM)

Item response theory analysis achieved a good fit after six iterations ($n = 321$). The discrimination parameters (α : 0.85–5.30, (> 0.700)) represented that the item response categories effectively differentiate between respondents with varying disability. The difficulty parameters (β : -4 – 4) show that a higher disabled patient would be more likely to have a 50% chance of certifying more severe response options than a patient with a lower disability. Participants had a 50% probability of selecting the categories β_1 (one or above), β_2 , β_3 and β_4 in D1.1, with difficulty parameters of -0.26 , -0.35 , 1.44 , and 2.85 , respectively. The difficulty level of the bold items shifted towards higher disability levels compared to the average level (the positive values) (Table 4).

The test characteristic curve (Fig. 2a) showed a shift towards higher levels of disability. The test information function curve shows function peak located on the right side of the point indicating an average level of disability in the studied population (point of zero on the X-axis) (Fig. 2b).

The student's t-test for independent samples represented insignificant gender difference in disability at the scale level. The analysis revealed that women ($n = 88$; mean = 53.63, SD = 30.82) did not report significantly higher disability than men ($n = 233$; mean = 44.74, SD = 34.26), $t(319) = -1.95$, $p > 0.05$. To test for b-DIF and ab-DIF, on 321 observations across 36 items, the grouping variable was sex. The ability of interest was the total score disability. Non-Uniform DIF was reported ($P < 0.05$) for items 5.1 and 6.8, indicating significant DIF in discrimination and difficulty between men and women. For all other items, gender difference was uniform (see Online Resource 3).

4 Discussion

The 36-item WHODAS 2.0 indicated robust translation validity and psychometric properties in application through phone survey among trauma patients.

The study demonstrated that psychometric characteristics are reflective of disability levels caused by health condition, rather than administration mode, within the population, although data collection methods, such as phone surveys, may introduce variations. According to IRT and CFA, the tool effectively differentiates disability severity level [50, 57, 62, 63]. Furthermore, the strong consistency, high reliability, and validity indices point toward WHODAS 2.0's robustness, regardless of administration mode. There are however potential reasons to conduct future studies comparing telephone administration mode with other common methods such as interviews [16, 18, 31, 50], including its remote data collection, less costs [64], better response rate [65], easier quality control [66], no need to travel to study site, problematic when interviewing with elderly patients and those with hearing problems [14], respondents socially desirable answers [67] and weak reliability and validity versus mail ones [68].

The data verified excellent consistency over time and showed better results comparing to a study in people with hand injuries (ICC = 0.88) [69]. Multiple studies identified high ICC in large populations [9], mentally disordered patients [70], and patients with sclerosis [17]. Similarly, we represented excellent stability reliability in trauma patients. The SEM of 1.43 for WHODAS 2.0 indicated that score variations within this range (± 1.96 SEM) with 95% confidence reflect measurement error rather than true changes in patients' disability and help clinicians recognize that changes smaller than the SEM are likely not clinically significant, while larger changes may indicate true improvements or declines. The study indicated that WHODAS 2.0 score changes exceeding 3.96 (SDC) likely reflect a real change in functional condition, rather than measurement error. It guides clinicians to identify meaningful changes in function during assessment or intervention. A study on 5809 people with multiple sclerosis reported 1.971 and 5.466 as values of SEM and SDC for 36-item WHODAS 2.0 [71]. The discrepancy may arise from differences in sample characteristics or methodologies, underscoring the need for context-specific thresholds. Future research should examine these metrics in diverse populations to improve the application of the WHODAS 2.0 and enhance patient care.

Our ideal internal consistency ($\alpha = 0.95$) was in consistent with studies in patients with sclerosis ($\alpha > 0.7$) [17], with hearing loss ($\alpha = 0.94$) [72], with spinal cord injury ($\alpha = 0.97$) [73], and Taiwanese population ($\alpha = 0.91$) [17] and others (α : 0.92, 0.94, 0.96) [9, 34, 74]. It represented better results compared with studies in Syrian refugees ($\alpha = 0.74$) [75], people with musculoskeletal pain ($\alpha = 0.84$) [76] and female population after pregnancy [77]. Furthermore, "items" under one factor measured the same constraint, confirmed CR [78], in line with studies on 298 adults receiving individual counselling (0.72–0.93) [79] and road traffic survivors [50].

The study contributes to building evidence for excellent validity at total and subscale levels among post-crash trauma survivors. The CFA model confirmed fit of the original six-facet model and its psychometric properties. The goodness of

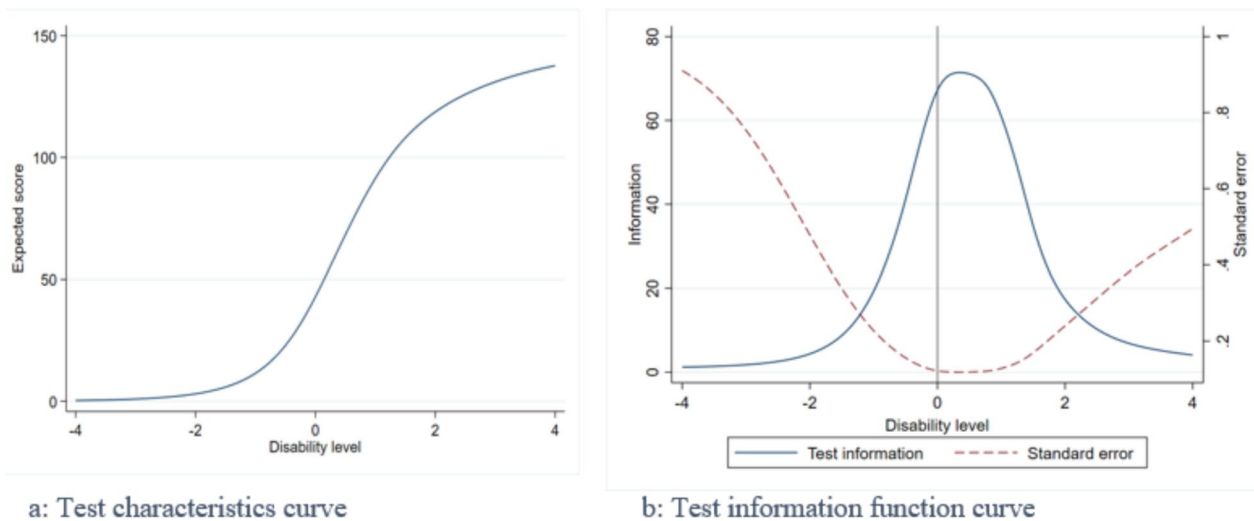


Fig. 2 Test curves for the WHODAS 2.0 in phone surveys

fit indices verified model's quality, similar to findings from a study on 1119 patients with chronic conditions in Europe [63] and other studies [50, 62]. Convergent validity was confirmed as variance explained by indicators exceeds that by error [12]. In opposite, a study of 298 adult patrons in therapy, reported AVEs < 0.5, rejecting convergent validity [79]. Item factor loadings ($p < 0.0001$) further supported convergent validity. The Fornell-larcker criterion and HTMT matrix supported discriminant validity.

Two items Eating (0.55) and Sexual activity (0.67) showed low factor loadings on their related factors. Similarly, some papers have reached similar findings [62, 77]. Some other research studies found similar findings on Chinese and Persian WHODAS 2.0 and indicated better fit with the reduced model [16, 62]. Moreover, some papers indicated the relevance of sexual activities when disease severity is evaluated by healthcare professionals as they consider sexual functions to be of relevant aspects [58]. The issue of sexual activities is related to patients' physical functioning or chronic diseases for instance Multiple Sclerosis [17]. Accordingly, Patients'socio-economic status, cultural background, and health-related factors influence how they perceive and report challenging conditions related to these items. The exclusion of these items must be approached cautiously because it could limit the generalizability of WHODAS 2.0 across diverse patient populations. As such, it is important to discuss their potential limitations in diverse clinical contexts. This could result in development of condition-based versions of WHODAS 2.0. Moreover, we highlight the importance of considering population-specific characteristics when we interpret disability assessments. The influence of administration mode on the emerging evidence should be approached cautiously, although video calls may offer a viable alternative to in-person data collection methods [80]. Addressing these points, help us provide a more nuanced interpretation and greater clarity.

The inter subscale correlation coefficients (0.7) support strong relationships between activities requiring physical movement: Mobility, Life activities, Self-care and Participation, consistent with a study of 600 women in different postpartum periods [77]. This is expected as physical activities are correlated. Furthermore, comparisons with previous studies are limited since, to our knowledge, this is the first application of WHODAS 2.0 in trauma survivors via phone survey.

Concerning IRT analysis, the items effectively differentiated respondents across ability levels, showing good discriminating power (> 0.700), and a direct relationship with the latent trait [57]. The highest discrimination value was 5.30 (D5.1) and the lowest value was 0.85 (D1.2), still above the recommended value. A shift towards higher disability scores indicates the tool is more sensitive at distinguishing between "much more" and "more" disabled respondents than between "more" and "less" disabled ones. A cohort study on patients with chronic musculoskeletal pain reported similar findings [81]. The items' difficulty identified a slight shift towards higher disability, namely, a participant should suffer from slightly worse disability (compared with the average population rate) to achieve fifty-fifty probability of giving an answer categorized as worse disability [81]. The items D6.5 and D6.6 were easiest for respondents at level one, while item D1.5 was the most difficult at an alternative level four. Methods such as Rasch analysis consider items' difficulty, while IRT estimates ability to discriminate different disability levels. Overall, WHODAS 2.0 performs well at determining disability level, particularly higher levels.

Regarding DIF, two items (D5.1, D6.8) were biased by sex and overestimate the disability level in men compared to women or vice versa, opposing some studies [8, 31]. Men and women at the same disability level responded similarly to 34 items and their group mean differences by sex were not attributable to gender-item bias. This contrasts with other studies due to differences in populations (patients with depression or acute myocardial infarction), major participants (male (69%) [82], female (67%) [8], ours' male (73%)), and methodologies (Partial Credit Model [82], kernel-smoothing [8], GRM in this work). Additionally, item D6.8 was not included in WHODAS 2.0 short version. A schizophrenia study also revealed no DIF by sex using Partial Credit Model [83].

Our sample includes a higher proportion of men (72%) and employed individuals (75%) among hospitalized traffic accident survivors. This gender distribution reflects the typical male dominance observed in road accident statistics [84, 85]. However, the situation is reversed in the case of fatal car crashes [86]. The composition of our sample, regarding gender and employment status, may reflect certain characteristics of trauma patients who are more likely to develop disabilities. With respect to the eligibility criteria, we applied broad criteria to capture a wide range of injury severities and outcomes. This allowed us to include participants with mild to moderate cognitive challenges. Then, we acknowledge that these sample characteristics may limit the generalizability of the results to certain populations, such as those who are not employed or who experience more severe cognitive impairments. There is a need for further research in more diverse trauma populations.

The Persian WHODAS 2.0 was developed upon WHO permission and translation guidelines. The available translations [16, 18], were not assessed in RTA trauma patients and were not developed in a public-friendly format. We needed a robust, informal Persian version to apply within the national post-crash PERSIAN Traffic Cohort Study platform. According to its excellent translation validity, it could be proposed for assessment in other languages and disability-related conditions.

4.1 Strengths and limitations

This study may be one of the first to evaluate the psychometric properties of the 36-item WHODAS 2.0 using phone interviews with traumatized adult patients, however, we acknowledge challenges associated with this approach. Although telephone interviews can be conducted more quickly and conveniently than a face-to-face interviews, respondents may be more inclined to give socially desirable answers, source of potentially distort findings [67]. This social desirability bias may affect the reliability of the data. Interviewer's specific information to clarify misunderstandings could further impact the responses [87]. Then, we recommend further research using varied valid data collection methods to mitigate potential bias.

The trauma patients' disability and functioning profile is determined at low cost and effort. The male dominance in the sample reflects typical gender proportion in road traffic accidents [88]. A limitation was the inclusion of only above 18-year-old patients due to feasibility issues. Then, similar studies among adolescents are recommended. The findings do not assess the tool's longitudinal performance post-trauma and further research is needed to evaluate its responsiveness over time. Furthermore, we should evaluate the tool among elderly and patients with hearing problem.

4.2 Conclusion

The validation of WHODAS 2.0 is realized within the post-crash PERSIAN Traffic Cohort platform, using the advanced methods via Smart PLS structural equation modelling, considering the acceptable levels for the factor analysis and the parameters of item-response theory. The factor analysis validates the latent constructs and IRT validates the discrimination power of the items. The Persian 36-item WHODAS 2.0, developed and validated by translation, showed good psychometric properties, good consistency over time, internal consistency reliability, and composite reliability, in trauma survivors in clinical and research settings. This tool is good at measuring disability levels, and mainly good at distinguishing higher levels. The study strengthens the existing literature demonstrating that WHODAS 2.0 indicated decent and reliable psychometric properties among the study population, and translation validity.

Acknowledgements The authors would like to thank the participants in the study.

Author contributions HSB conceived and designed the study. HSB, MAJ and NS conducted the analyses, and drafted the manuscript. HSB, NS, SBH, FP, and MG interpreted the results. NS and MG involved in execution of the study. SBH, MAJ, NS, SBH, FP, and MG contributed in the critical drafting and revising of the manuscript for important intellectual content. All authors read and approved the final manuscript to be published.

Data availability The data could be available by a request by sending email to the corresponding author Dr. Homayoun Sadeghi-Bazargani.

Declarations

Ethics approval and consent to participate This study appreciated ethical principles of the declaration of Helsinki. The road traffic injury research center ethics committee, and regional ethics committee, Tabriz University of Medical Sciences approved the research project (PERSIAN Traffic Cohort Study ethics code: IR.TBZMED.REC.1398.543; PhD thesis ethics code: ir.tbzmed.rec.1399.971, in 25 January 2021). The study was done within the PhD research thesis. An informed consent was required from all participants. Their anonymity were preserved.

Competing interests The authors declare no competing interests.

Consent for publication All authors read the work and announced their consent for its publication.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. World Health Organization. *Disability*. 7 March 2023 [cited 2023; Available from: <https://www.who.int/news-room/fact-sheets/detail/disability-and-health>.
2. United Nations, D.o.E.a.S.A.D. *Disability*. Factsheet on Persons with Disabilities, 2023 2023 [cited 2023; Available from: <https://www.un.org/development/desa/disabilities/resources/factsheet-on-persons-with-disabilities.html>.
3. Krug EG, Sharma GK, Lozano R. The global burden of injuries. *Am J Public Health*. 2000;90(4):523–6.
4. Sehat Z, et al. Disability and post-trauma stress in the population over 15 years old in Kashan, Iran: a population-based study. *Chin J Traumatol*. 2020;23(6):351–5.
5. Sadeghi-Bazargani H, et al. Time to fast-track interventions to road traffic crises in Iran. *The Lancet*. 2024;403(10442):2373–4.
6. Vakili M, et al. The burden of road traffic injuries in Yazd Province—Iran. *Bull Emerg Trauma*. 2016;4(4):216–22.
7. van der Ploeg HP, et al. Physical activity for people with a disability. *Sports Med*. 2004;34(10):639–49.
8. Luciano JV, et al. The 12-item world health organization disability assessment schedule II (WHO-DAS II): a nonparametric item response analysis. *BMC Med Res Methodol*. 2010;10(1):45.
9. Üstün TB, et al. Developing the world health organization disability assessment schedule 2.0. *Bull World Health Organ*. 2010;88(11):815–23.
10. World Health Organization. *WHO Disability Assessment Schedule 2.0 (WHODAS 2.0)*. 2018; Available from: <https://www.who.int/classifications/icf/whodasii/en/>, retrieved at 29.03.2020.
11. Papadakaki M, et al. Psychological distress and physical disability in patients sustaining severe injuries in road traffic crashes: Results from a one-year cohort study from three European countries. *Injury*. 2017;48(2):297–306.
12. Ćwirlej-Sozańska A, et al. Psychometric properties and validation of the polish version of the 12-item WHODAS 2.0. *BMC Public Health*. 2020;20(1):1203.
13. T. Bedirhan Üstün, et al., *Chapter 58: The World Health Surveys, in Health Systems Performance Assessment*. 2003, World Health Organization: Geneva. p. 797–808.
14. Shahedifar N, et al. Psychometric properties of the 12-item WHODAS applied through phone survey: an experience in PERSIAN Traffic Cohort. *Health Qual Life Outcomes*. 2022;20(1):106.
15. Norouzi H, et al. Psychometric properties of the world health organization disability assessment scale 2.0 among Iranian cancer patients. *Middle East Journal of Cancer*. 2020;11(3):333–42.
16. Rajezi-fahani S, et al. Validity of the 36-item Persian (Farsi) version of the world health organization disability assessment schedule (WHODAS) 2.0. *Int J Mental Health*. 2019;48(1):14–39.
17. Salehi R, et al. Validity and reliability of the world health organization disability assessment schedule 2.0 36-item persian version for persons with multiple sclerosis. *Korean J Fam Med*. 2020;41(3):195–201.
18. Abedzadeh-kalahroudi, M., et al., *Psychometric properties of the world health organization disability assessment schedule II -12 Item (WHODAS II) in trauma patients*. *Injury*, 2016. **47**(5): p. 1104–1108.
19. Marin S, et al. The protocol for validating phone interview tools on post-discharge consequences of road traffic injuries. *J Inj Violence Res*. 2020;12(3):45–51.
20. Roshanfekar P, et al. Prevalence and socio-economic determinants of disabilities caused by road traffic accidents in Iran; A national survey. *Bull Emergency and Trauma*. 2019;7:60–6.
21. Sadeghi-Bazargani H, et al. PERSIAN Traffic Safety and Health Cohort: a study protocol on postcrash mental and physical health consequences. *Inj Prev*. 2022;28(3):269–79.
22. Marin S, et al. The protocol for validating phone interview tools on post-discharge consequences of road traffic injuries. *J Inj Violence Res*. 2020. <https://doi.org/10.5249/jivr.v12i3.1368>.

23. Steffen T, Seney M. Test-retest reliability and minimal detectable change on balance and ambulation tests, the 36-item short-form health survey, and the unified Parkinson disease rating scale in people with parkinsonism. *Phys Ther.* 2008;88(6):733–46.
24. Amick R, Chaparro A, Patterson J. Test-retest reliability of the sway balance mobile application. *J Mobile Tech Med.* 2015;4:40–7.
25. StataCorp, *Stata Item Response Theory Reference Manual Release 17.* 2021, College Station, TX: StataCorp LLC.
26. Nikan F, et al. Designation and psychometric properties of the short form postpartum quality of life questionnaire (SF-PQOL): an application of multidimensional item response theory and genetic algorithm. *Health Promot Perspect.* 2018;8(3):215–24.
27. Rathbone, A., S. Shaw, and D. Kumbhare, *ICC. sample. size: calculation of sample size and power for ICC.* R Package Version, 2015. **1.**
28. Chin WW, Marcolin BL, Newsted PR. A partial least squares latent variable modeling approach for measuring interaction effects: results from a monte carlo simulation study and an electronic-mail emotion/adoption study. *Inf Syst Res.* 2003;14(2):189–217.
29. Soper, D.S., *A-priori Sample Size Calculator for Structural Equation Models [Software].* 2023.
30. World Health Organization. *International Classification of Functioning, Disability and Health (ICF).* 2021 [cited 2021 24/05/2021]; Available from: <https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health>.
31. Kirchberger I, et al. Feasibility and psychometric properties of the German 12-item WHO Disability Assessment Schedule (WHODAS 2.0) in a population-based sample of patients with myocardial infarction from the MONICA/KORA myocardial infarction registry. *Popul Health Metrics.* 2014;12(1):27.
32. Ustun TB, et al. *Measuring health and disability manual for WHO Disability Assessment Schedule.* Geneva: World Health Organization; 2010.
33. Silva AG, et al. Self-reported disability: association with lower extremity performance and other determinants in older adults attending primary care. *Phys Ther.* 2015;95(12):1628–37.
34. Irfan, M., et al., *TRANSLATION AND VALIDATION IN PASHTO (3): WORLD HEALTH ORGANIZATION DISABILITY ASSESSMENT SCHEDULE 2.0.* Journal of Postgraduate Medical Institute, 2017. **31**(4).
35. StataCorp, *Stata statistical software: release 14.* 2017, Statacorp Ip College Station, TX.
36. Ringle, C.M., S. Wende, and B. J.-M., *SmartPLS 3.* Boenningstedt: SmartPLS GmbH, <http://www.smartpls.com>, 2015.
37. Halvorsrud, L. *Measuring Quality of Life among Older Adults. : Validation of the Norwegian WHOQOL-Old.* 2010.
38. Laghousi D, et al. Validity of the physical activity scale for the elderly: results from the neyshabur longitudinal study on ageing. *Med J Tabriz Univ Med Sci.* 2024;46:407–18.
39. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–63.
40. Chengalur-Smith I, Duchessi P, Gil-Garcia JR. Information sharing and business systems leveraging in supply chains: an empirical investigation of one web-based application. *Inf Manag.* 2012;49(1):58–67.
41. Terwee CB, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34–42.
42. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep.* 1966;19(1):3–11.
43. Vejlggaard C, et al. Reliability and smallest detectable change of the danish version of the pain self-efficacy questionnaire in patients with chronic low back pain. *Scand J Pain.* 2021;21(4):809–13.
44. He S-J, et al. Validation of an 8-item recovery assessment scale (RAS-8) for people with schizophrenia in China. *Health Qual Life Outcomes.* 2021;19(1):119.
45. Ghorbani Vajargah P, et al. Psychometric evaluation of the heart failure somatic perception scale in Iranian heart failure patients: a cross-sectional study. *Ann Med Surg (Lond).* 2023;85(11):5396–402.
46. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod.* 2013;38(1):52–4.
47. Sovey S, et al. Exploratory and confirmatory factor analysis for disposition levels of computational thinking instrument among secondary school students. *Başlık.* 2022. <https://doi.org/10.12973/eu-jer.11.2.639>.
48. Villalustre L, Cueli M. Assessing the computational thinking of pre-service teachers: a gender and robotics programming experience analysis. *Education Sciences.* 2023;13(10):1032.
49. Sarstedt M, Ringle CM, Hair JF. Partial Least Squares Structural Equation Modeling. In: Homburg C, Klarmann M, Vomberg AE, editors. *Handbook of Market Research.* Cham: Springer International Publishing; 2020. p. 1–47.
50. Denu ZA, et al. The 12 items Amharic version WHODAS-2 showed cultural adaptation and used to measure disability among road traffic trauma victims in Ethiopia. *BMC Psychol.* 2021;9(1):1.
51. Bido D, Silva D. Smartpls 3: specification, estimation, evaluation and reporting. *Administração Ensino e Pesquisa.* 2019;20:465–513.
52. Hair J, et al. An updated and expanded assessment of PLS-SEM in information systems research. *Ind Manag Data Syst.* 2017;117(3):442–58.
53. Cohen, J., *Statistical power analysis for the behavioral sciences.* 2nd ed. Vol. 2nd Edition. 1988, New York: Routledge.
54. Roldán JL, Sánchez-Franco MJ, et al. Variance-Based Structural Equation Modeling: Guidelines for Using Partial Least Squares in Information Systems Research. In: Mora M, et al., editors. *Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems.* Hershey: IGI Global; 2012. p. 193–221.
55. Latan, H. and I. Ghozali, *Partial Least Squares: Concept, Technique and Application using Program SmartPLS for Empirical Research.* BP UNDIP, 2012.
56. Nguyen TH, et al. An introduction to item response theory for patient-reported outcome measurement. *Patient.* 2014;7(1):23–35.
57. Tezza R, Bornia AC, de Andrade DF. Measuring web usability using item response theory: principles, features and opportunities☆. *Interact Comput.* 2011;23(2):167–75.
58. Berrio A, Gómez-Benito J, Guilera G. Differential item functioning in the WHODAS 2.0 scale in schizophrenia: an application of the rasch trees method based on demographic and clinical covariates. *Assessment.* 2021. <https://doi.org/10.1177/10731911211036746>.
59. Crane PK, van Belle G, Larson EB. Test bias in a cognitive test: differential item functioning in the CASI. *Stat Med.* 2004;23(2):241–56.
60. Norén P, et al. Psychometric evaluation of the WHODAS 2.0 and prevalence of disability in a Swedish general population. *J Patient-Rep Outcomes.* 2023;7(1):36.

61. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Eq Model Multidisciplinary J.* 1999;6(1):1–55.
62. Chiu TY, et al. 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 Dev Disabil.* 2014;35(11):2812–20.
63. Garin O, et al. Validation of the “World Health Organization Disability Assessment Schedule, WHODAS-2” in patients with chronic diseases. *Health Qual Life Outcomes.* 2010;8(1):51.
64. Arita S, et al. Use of interviewer-administered telephone surveys during infectious disease outbreaks, epidemics and pandemics: a scoping review. *BMJ Glob Health.* 2023;8(5): e011109.
65. Jones T, Baxter M, Khanduja V. A quick guide to survey research. *The Annals of The Royal College of Surgeons of England.* 2013;95(1):5–7.
66. Kempf AM, Remington PL. New challenges for telephone survey research in the twenty-first century. *Annu Rev Public Health.* 2007;28:113–26.
67. Holbrook A, Green M, Krosnick J. Telephone versus face-to-face interviewing of national probability samples with long questionnaires: comparisons of respondent satisficing and social desirability response bias. *Public Opin Q.* 2003. <https://doi.org/10.1086/346010>.
68. Erhart M, et al. Effects of phone versus mail survey methods on the measurement of health-related quality of life and emotional and behavioural problems in adolescents. *BMC Public Health.* 2009;9:491.
69. Marom BS, et al. Cross-cultural adaptation of the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) for Hebrew-speaking subjects with and without hand injury. *Disabil Rehabil.* 2017;39(12):1155–61.
70. Svanborg C, et al. Evidence for validity of the Swedish self-rated 36-item version of the world health organization disability assessment schedule 2.0 (WHODAS 2.0) in patients with mental disorders: a multi-centre cross-sectional study using Rasch analysis. *J Patient-Rep Outcomes.* 2022;6(1):45.
71. Young CA, et al. Measuring disability in multiple sclerosis: the WHODAS 2.0. *Qual Life Res.* 2023;32(11):3235–46.
72. Chisolm TH, et al. The WHO-DAS II: psychometric properties in the measurement of functional health status in adults with acquired hearing loss. *Trends Amplif.* 2005;9(3):111–26.
73. Wolf AC, et al. The world health organization disability assessment scale, WHODAS II: reliability and validity in the measurement of activity and participation in a spinal cord injury population. *J Rehabil Med.* 2012;44(9):747–55.
74. Bejer A, et al. Psychometric properties of the Polish version of the 36-item WHODAS 2.0 in patients with hip and knee osteoarthritis. *Qual Life Res.* 2021;30(8):2415–27.
75. Akhtar A, et al. Exploratory and confirmatory factor analysis of the 12-item Arabic world health organization disability assessment schedule (WHODAS 20) as a screening tool for Syrian refugees. *BJPsych Open.* 2021;7(6):e190.
76. Silva C, et al. Adaptation and validation of WHODAS 2.0 in patients with musculoskeletal pain. *Rev Saude Publica.* 2013;47(4):752–8.
77. Silveira C, et al. Validation of the 36-item version of the WHO Disability Assessment Schedule 2.0 (WHODAS 2.0) for assessing women's disability and functioning associated with maternal morbidity. *Revista Brasileira de Ginecologia e Obstetricia / RBGO Gynecology and Obstetrics.* 2017. <https://doi.org/10.1055/s-0037-1598599>.
78. Esposito Vinzi, V., et al., *Handbook of Partial Least Squares: Concepts, Methods and Applications.* 2010.
79. Tabet SM, et al. An analysis of the world health organization disability assessment schedule 2.0 measurement model using partial least squares-structural equation modeling. *Assessment.* 2020;27(8):1731–47.
80. Gray D, et al. Development, reliability, validity, and acceptability of the remote administration of the Edinburgh Cognitive and Behavioural ALS Screen (ECAS). *Amyotroph Lateral Scler Frontotemporal Degener.* 2024;25(1–2):96–103.
81. Saltychev M, et al. A study of the psychometric properties of 12-item world health organization disability assessment schedule 2.0 in a large population of people with chronic musculoskeletal pain. *Clin Rehabil.* 2017;31(2):262–72.
82. Kirchberger I, et al. Feasibility and psychometric properties of the German 12-item WHO Disability Assessment Schedule (WHODAS 2.0) in a population-based sample of patients with myocardial infarction from the MONICA/KORA myocardial infarction registry. *Popul Health Metrics.* 2014;12:27.
83. Berrío Á, et al. Differential Item Functioning in the WHODAS 2.0 Scale in Schizophrenia: an application of the rasch trees method based on demographic and clinical covariates. *Assessment.* 2021. <https://doi.org/10.1177/10731911211036746>.
84. Al-Balbissi AH. Role of gender in road accidents. *Traffic Inj Prev.* 2003;4(1):64–73.
85. Bener A, Crundall D. Role of gender and driver behaviour in road traffic crashes. *Int J Crashworthiness.* 2008;13:331–6.
86. Abrams MZ, Bass CR. Female vs. male relative fatality risk in fatal motor vehicle crashes in the US, 1975–2020. *PLoS ONE.* 2024;19(2):e0297211.
87. Bispo Júnior JP. Social desirability bias in qualitative health research. *Rev Saude Publica.* 2022;56:101.
88. Sadeghi-Bazargani H, et al. Epidemiology of road traffic injury fatalities among car users; a study based on forensic medicine data in East Azerbaijan of Iran. *Bull Emergency Trauma.* 2018;6:146–54.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.