Rasch Model Analysis of the Indonesian Version of World Health Organization Disability Assessment Schedule (WHODAS 2.0)

Lukman Hilfi¹,², Nur Atik³, Ardini Saptaningsih Raksanagara², Deni Kurniadi Sunjaya², Sekar Ayu Paramita¹,², Chiho Yamazaki¹, Hiroshi Koyama¹ and Kei Hamazaki¹

¹ Department of Public Health, Gunma University Graduate School of Medicine, 3-39-22 Showa-machi, Maebashi, Gunma 371-8511, Japan
² Department of Public Health, Faculty of Medicine, Universitas Padjadjaran, Bandung, Indonesia
³ Department of Biomedical Sciences, Faculty of Medicine, Universitas Padjadjaran, Bandung, Indonesia

Abstract
Background & Aims: This study evaluated the construct validity of the Indonesian version of WHODAS 2.0 among persons with physical impairment.

Methods: WHODAS 2.0 was self-administered to 212 participants with physical impairment in Bandung, Indonesia. The Rasch model was used to analyse the instrument’s construct validity.

Results: Data of 212 participants were analysed. The average outfit mean-square (MnSq) of the 36-item and 32-item versions of WHODAS 2.0 satisfied the Rasch model expectations (0.99±0.28 vs 0.99±0.24). Item D2.5 ‘walking a long distance such as a kilometre’ and item D4.5 ‘sexual activity’ were identified as misfitting items in both versions (infit or outfit MnSq >1.4). The variance explained by measures of the 32-item version was 56.7%, and the 36-item version was 49.0%. Both versions’ reliability and separation index were excellent, with Cronbach’s alpha >0.90 and a separation index >2. The response category function and targeting did not fully satisfy the Rasch model expectations. A strong correlation between both versions of WHODAS 2.0 and WHOQOL BREF (r>0.60) established the convergent validity.

Conclusions: The Indonesian version of the 32-item version of WHODAS 2.0 has acceptable construct validity in a physical impairment sample in Bandung, Indonesia.

Introduction
Physical impairments are problems such as a significant deviation or loss of body function or structure.¹ Disability is the umbrella term for impairments, activity limitations and participation restrictions, and refers to the negative aspects of the interaction between an individual (with a health condition) and that individual’s contextual factors (environmental and personal factors).² A person with disabilities frequently experiences physical, social and environmental barriers limiting their activity and participation in the community.³,⁴

The United Nations Sustainable Development Goals (SDGs) 2015-2030⁵ and the Convention on the Rights of Persons with Disabilities (CRPD) urge countries to collect accurate data on disability to understand the impact of barriers on persons with disabilities.⁶,⁷ To address the data accuracy on disability, the International Classification of Functioning, Disability and Health (ICF) framework on disability is broadly used.² This framework redefines disability as a limitation on activity and a restriction on participation.⁸ Reflecting the ICF framework, the World Health Organization Disability Assessment Schedule (WHODAS 2.0) has been developed to assess the activity limitations and participation restrictions of persons with disabilities.⁴

WHODAS 2.0 was translated into many languages and developed specifically to measure activities and participation.⁴ The Indonesian version of the WHODAS 2.0 has been used in several censuses to assess the preva-
lence rate of disability. However, a lack of evidence on the accuracy of WHODAS 2.0 has become an issue in Indonesia. To provide an accurate measurement, critical psychometric properties of a measurement instrument, such as reliability and validity, must be exhibited. The psychometric theory offers two approaches to analyse an instrument to determine the validity and reliability: the factor analysis based on a classical test theory (CTT) approach and the Rasch model analysis based on an item response theory (IRT) approach. The CTT analysis focuses mainly on the observed total score of analysing validity and reliability. Using this approach, the WHODAS 2.0 was found valid in internal diseases, musculoskeletal conditions, stroke and back pain. A study conducted by Yuliana et al. (2020) also concluded that the WHODAS 2.0 was valid and reliable to assess the level of disability of stroke survivors in Indonesia.

However, the CTT has some conceptual limitations. The CTT often uses the ordinal response scale as an interval scale and assumes that a sum of the score observed can define a person’s actual ability or trait. However, it is not appropriate for measuring empirical reality and can lead to erroneous conclusions because the test score achieved by a person often differs from a person’s actual score. Moreover, this approach does not consider the ordered continuum of items and lacks the response category data additivity.

Rasch model analysis provides supplementary evidence for the construct validity of the instrument and focuses on the probabilistic unidimensional analysis that models the interaction between the item and person. This approach can distribute the persons’ ability and the items’ difficulty levels along the same logit scale. Rasch model analysis gives statistical information on the ordinal response categories’ appropriateness and converts ordinal scores into equal-interval level data to increase measurement precision.

To date, WHODAS 2.0 has been evaluated using Rasch model analysis. In previous studies, Rasch model analysis proved that WHODAS 2.0 has acceptable validity and reliability to assess several health conditions and cultural environments. In contrast, a study by Kutlay et al. (2009), conducted in Turkey, found that the construct validity of WHODAS 2.0 in patients with osteoarthritis, for some domains, was less than satisfactory.

The Indonesian version of WHODAS 2.0 has never been psychometrically calibrated and validated using Rasch model analysis. Therefore, this study was conducted to provide evidence of construct validity of the WHODAS 2.0 among persons with physical impairment using the Rasch model analysis approach. The results of this study are essential to assess the value of WHODAS 2.0 and its utility in Indonesia.

Materials and Methods

This study was conducted in April-May 2019 using self-administered WHODAS 2.0. Ethics approval for the study has been granted from the Research Ethical Committee of Universitas Padjadjaran (No. 140/UN6.KEP/EC/2019) and Gunma University (No. HS2019-318). The Social Office of Bandung City has approved this study permission (No. 460/1966-Dinsosangkis). Prior to data collection, trained surveyors explained the study and received the participants’ informed consent. The participants filled in the questionnaire by themselves but could ask questions to the surveyor if they did not understand. The participants could withdraw from the study at any stage.

Study Setting

This study was carried out in Bandung, Indonesia. Bandung is a large city with an area of 168 km² and serves as the capital of the West Java Province. Bandung is the third most populous city in Indonesia after Jakarta (the country’s capital) and Surabaya and is one of the most important National Activity Centres for political, economic, and social activities. The industry sector is critical since it contributes significantly to Bandung’s gross domestic product (GDP), utilises local labour, and stimulates local economic activity.

Bandung’s population is estimated to be over 2.5 million people, expanding at a rate of 3.5% per year, and is expected to reach 4.1 million people by 2031, much beyond the city’s current capacity. In 2019, the Bandung Social and Welfare Office registered 5358 people as having impairments, with 2344 of them suffering from physical impairments or a combination of physical and other impairments. A total of 1380 (59%) had only a physical impairment.

Participants

We recruited participants with a physical impairment. We used two steps for the sampling technique: purposive sampling and stratified random sampling. Purposive sampling was performed to collect data from potential participants aged 18 and 60 years who had only a physical impairment, had no cognitive decline and lived in Bandung. Persons with other impairments, such as vision impairment, hearing impairment or cognitive impairment, were not eligible for this study. A list of 1380 persons with a physical impairment was obtained from the social and welfare office of Bandung.

We then used stratified random sampling to achieve the targeted sample size within 121 to 303. We calculated a maximum sample size of 303 based on an estimated 22% of persons with disabilities in Indonesia. The minimum sample size of 121 was estimated using EZR version 1.51 (the standard deviation of 14, a confidence interval of 5, and a confidence level of 95%). We had 271 participants with physical impairments who responded to the questionnaire.

Instruments

(1) Demography Data

Using a questionnaire, we gathered socio-demographic information such as age, gender, married status, education status, occupation status, type of impairment, and assistive device usage status.
(2) WHODAS 2.0
The World Health Organization Disability Assessment Schedule (WHODAS 2.0) is a standardised method for measuring health and disability across cultures developed by the WHO.\(^4\) WHODAS 2.0 captures the level of functioning during the 30 days preceding the interview in six domains of life. We used the Indonesian version of the self-administered WHODAS 2.0, which was officially obtained from the WHODAS group. It has been pilottested, back-translated into English, and used in previous studies.\(^31,37-39\)

The WHODAS 2.0 items are designed on 5 Likert scales from 0 to 4, where 0 is no difficulty; 1 is mild difficulty; 2 is moderate difficulty; 3 is severe difficulty, and 4 is extreme difficulty/cannot do. WHODAS 2.0 generates scores for the six domains of functioning (cognitive, getting around, self-care, getting along with people, daily activities, and participation in society) and calculates an overall functioning score. The score ranges from 0 to 100, in which higher scores indicate more significant disability (0 = no disability; 100 = very high degree of disability). The WHODAS 2.0 total score for the worker or student is the sum of 36 items. For those who were not working or in school, the score is the sum of 32 items. The 36-item version of WHODAS 2.0 was used in this study to measure the level of disability of the participants. Following the WHODAS 2.0 manual,\(^4\) participants with more than 30% missing values were excluded from the analysis.

(3) The Abbreviated Version of World Health Organization Quality of Life Instrument
The abbreviated version of World Health Organization Quality of Life Instrument (WHOQOL-BREF) is used to analyse persons’ quality of life (QOL). As in the study of Üstün et al. (2010), this study also used WHOQOL-BREF to measure the convergent validity of WHODAS 2.0.\(^30\) This instrument has 26 items comprising 2 items measuring overall QOL and general health, and 24 items measuring four domains: physical, psychological, social relationships, and environmental.\(^40\) The Indonesian version of the WHOQOL-BREF is available and has been proven to be a valid questionnaire in Indonesia.\(^41,42\) In this study, Rasch model analysis showed that this instrument was unidimensional. Thus, the interval logit scale of WHOQOL-BREF was used in the analysis of WHODAS 2.0 convergent validity.

Statistical Analysis
We used Rasch model analysis to evaluate the construct validity of WHODAS 2.0. Rasch model analysis is a mathematical model which assumes that each scale item has a different value or difficulty level. It can provide item and person distributions based on the item’s difficulty and the person’s level of disability using a unit of measurement called a logit (log-odds unit). Rasch model analysis can convert raw ordinal scores into interval logit measures\(^25\) to satisfy the data assumption of normality.

According to Messick (1989), construct validity is the integration of any evidence that bears on the interpretation or meaning of test scores. Construct validity emphasizes the development of models explaining processes underlying performance on various tests and their relationships to other phenomena.\(^43\) In this study, we separately evaluated the construct validity of the 36-item version, 32-item version, and six domains of WHODAS 2.0 based on the following seven indicators.\(^14,44\)

(1) Fit Statistics
Fit statistics were used to evaluate the fit of items to the model.\(^45\) WHODAS 2.0 fit statistics were evaluated using the average outfit mean-square (MnSq) and the item and person’s fit statistics.\(^25,46\) MnSq is the mean of the squared residuals for an item, has a chi-square distribution and an expected value.\(^45\) Item fit statistics evaluate the performance of a given item in a test.\(^47\) If fit statistics are greater than the expected value, they suggest more distortion of the measurement system.\(^48\) A misfitting item suggests that the item measures a different and irrelevant construct. Person fit statistics indicate the degree to which persons’ response patterns follow what would be expected by the model.\(^49\)

There were three steps for fit statistical analysis\(^25\) used in this study.

1. The average outfit MnSq was evaluated to determine if the set of WHODAS 2.0 fit the Rasch model expectation. The MnSq $1.4$ logit is the cut-off point of the acceptable fit statistics,\(^25\) indicating 40% more variation than predicted by the Rasch model.\(^45\) Item infit or outfit MnSq $> 1.4$ logit was used to determine the misfitting item. Infit (inlier-sensitive or information-weighted fit) is relatively affected by unexpected responses closer to item and person measures. Outfit (outlier-sensitive fit) is relatively affected by unexpected responses distance from the item or person measures.\(^46\)

2. If items were misfit, participants with more than 30% missing values\(^30\) and infit or outfit MnSq $> 1.4$ logit (misfitting person) were excluded.

3. The remaining data were used to re-evaluate the average outfit MnSq of a set of items and the infit or outfit MnSq of each item to identify the misfitting item. We might not exclude the misfitting item if the other construct validity indicators satisfied the Rasch model expectations.

(2) Unidimensionality
Unidimensionality is a fundamental requirement of construct validity and is necessary to validate a single score of an instrument\(^10\) such as WHODAS 2.0. The unidimensionality assumes that the measurement instrument must measure only one trait at a time.\(^45\) A principal component analysis (PCA) of residuals evaluates the unidimensional exploration of additional explanatory dimensions in the data. The criterion for unidimensionality used in this study is that the first latent dimension should explain at least 50% of the total variance.\(^48\) If the measurement is unidimensional, a single score for the whole test can be used. In contrast, separate scores for each dimension should be considered if the instrument is not unidimensional.\(^51\)
(3) **Response Category Function**

The response category function assumes that the transition from one ordered response category to the next do not vary across the items. The response category function was evaluated according to the following four indicators.48,52

1. The average measures and the Andrich thresholds for each step category should advance monotonically.
2. The Andrich threshold difference should be between 1.4-5 logit.
3. The outfit MnSq for each step category should be less than 2.0.
4. Each category should have the same peak as the most probable response at some point along the continuum.

If the response category scale is disordered, we could adjust response options by collapsing the adjacent categories. However, as long as the instrument’s construct validity fits the Rasch model expectations, adjusting the response category is unnecessary.48

(4) **Reliability and Separation**

Reliability or internal consistency refers to how all the items or questions assess the same characteristic. Cronbach’s alpha, the average of all the correlations between each item and the total score, is often calculated to determine the internal consistency.53 The Cronbach’s alpha values range from 0 to 1.0. The closer the value is to 1.0, the less the variability can be attributed to measurement error.54 Any Cronbach’s alpha value between 0.80 and 0.90 is considered good, and a value of more than 0.90 is considered excellent.45

The separation index was also calculated to determine whether the instrument could differentiate items and persons with different difficulty levels. For differentiating at least two distinct groups, a separation index of 2 is required.25 When the instrument fails to meet the item and person separation criteria, it is likely because the items are not sufficiently different in terms of difficulty levels and do not match all the disability levels represented among the participants.55

(5) **Appropriateness of the item difficulty level for the sample**

The appropriateness of the item difficulty level for the sample was examined using targeting and can be seen on the Wright map (person-item map) in Fig. 2. On the left side of the map, the participants were ranked based on their latent traits (or in our study, their level of disability), and on the right side of the map, the items were ranked based on their difficulty (or in our study, the level of disability implied by the item). According to the Rasch model principle,56 we assumed that participants with lower disability levels had a higher probability of easily doing more difficult activities (more negative logits). Participants with higher disability levels were highly likely to find it difficult to do even the easier activities (more positive logits).

Rasch model analysis examines the targeting by evaluating the person-item mean difference and sample target coverage.57 A difference of zero and below one logit indicate a perfect targeting and a well-targeting of the scale, respectively, whereas a difference of more than one logit indicates mistargeting.58 The other targeting evaluation is to measure the sample target coverage. The more the scale covers the sample targets (targeting area), the better the targeting.59

(6) **Differential Item Functioning of Misfitting Item**

Differential item functioning (DIF) occurs when participants who are similar in certain latent traits but differ in another characteristic (DIF group) are more likely to score higher (or lower) on particular items than members of another group.54,49 A DIF is considered present if the average item calibration for the DIF group differs by >0.5 logit (p ≤ 0.05). In this study, DIF was explicitly evaluated to investigate the possible bias of the socio-demographic variables as the causes of the misfitting items. We evaluated the DIF group concerning socio-demographic variables, including age, sex, marital status, occupational status, and assistive device usage status.

(7) **Convergent Validity**

Convergent validity investigates how scores obtained from one test correlate with scores from another test.53 In this study, the WHOQOL-BREF was chosen for the convergent validity assessment because of the similarity of the questionnaire structure with the WHODAS 2.0. The WHODAS 2.0 measures functioning, while the WHOQOL measures subjective well-being.60,61 In addition, since the WHOQOL-BREF was unidimensional, its raw scores were converted to interval logit measures to evaluate Pearson’s correlation coefficient with that of the WHODAS 2.0 and, therefore, the convergent validity of the instruments was determined. A correlation coefficient (r) of around 0.60 was expected.62

**Ordinal to Interval Scale Conversion**

The conversion was used to interpret the size of the disability scores. The interval scale is used to analyse and compare outcomes across groups.50 If the instrument’s construct validity is confirmed, conversion of raw ordinal scores to interval logit measures can be applied.53 This will be helpful for WHODAS 2.0 users such as clinicians or rehabilitation staff69 who are not familiar with Rasch model analysis to use the interval logit measures directly.

In this study, the IBM SPSS statistics 27 (IBM Corp., Armonk, New York) was used for data entry, descriptive statistics, and correlation statistics, and the WINSTEPS 3.75 (Winsteps®, Beaverton, Oregon)49 was used to conduct the Rasch model analysis.14

**Results**

(1) **Fit Statistics**

A total of 271 participants with physical impairments was first used to evaluate the 36-item version, 32-item version and six domains of WHODAS 2.0. These were the steps to evaluate the fit statistics.

1. We evaluated the average outfit MnSq of the items
and identified misfitting items with infit or outfit MnSq > 1.4 logit using the 271 participants. The average outfit MnSq of the 36-item and 32-item versions of WHODAS 2.0 were 1.04 and 1.02, respectively. Misfitting items were found in the 36-item (D2.1, D2.2, D2.5, and D6.7) and 32-item (D4.5, D5.4, D6.1, and D6.7) versions.

2. According to the results of the fit statistics, 9 participants with more than 30% missing values were excluded. Since the fit statistic results using 262 remaining data did not satisfy the Rasch model expectations, 50 participants with infit or outfit MnSq > 1.4 logit (misfitting persons) were then excluded from the data set.

3. The remaining 212 participants were then finally used to re-evaluate the item fit statistics of the 36-item and 32-item versions of WHODAS 2.0. The average outfit MnSq improved to 0.99. The misfitting items of both versions were reduced to two, item D2.5 ‘walking a long distance such as a kilometre [or equivalent]’ (in 36-item version: infit MnSq = 1.88 and outfit MnSq = 1.85; in 32-item version: infit MnSq = 1.94 and outfit MnSq = 1.54) and item D4.5 ‘sexual activity’ (in 36-item version: infit MnSq = 1.80 and outfit MnSq = 1.44; in 32-item version: infit MnSq = 1.46).

These three steps were also used to separately evaluate the six domains of WHODAS 2.0. We found that item D3.4 (infit MnSq = 1.59; outfit MnSq = 1.76) and item D6.1 (infit MnSq = 2.99; outfit MnSq = 2.75) were misfitting items when we separately analysed domain 3 ‘self-care’ and domain 4 “getting along with people”, respectively. The construct validity evaluation of the 36-item, 32-item, and the domains of WHODAS 2.0 using 212 participants is shown in Table 1.

The demographic data of the 212 participants are shown in Table 2. The age of participants ranged from 18 to 60 years old, with a mean age of 41.1 ± 12.2 years. Among 212 participants, 58% were male, 52.8% were single, 85.9% graduated from elementary school or above, and 68.9% were non-worker; 50.9% of participants suffered from hemiplegia/paresis, and 51.7% of participants used assistive devices in their daily activities.

Table 3 shows the infit and outfit MnSq values of the WHODAS 2.0 using 212 remaining persons. Item D2.5, D3.4, and D4.5 were identified as misfitting items.

(2) **Unidimensionality**

The principal components analysis (PCA) of the Rasch analysis supported the unidimensionality of WHODAS 2.0. Table 1 shows the 49.0% variance explained by the Rasch dimension for the 36-item, which was very close to the desired 50% to fit the model. The variance explained by the measures for the 32-item versions was 56.3%, supporting the unidimensionality, fitting the model and reporting good precision for the physical impairment sample.

(3) **Response Category Function**

As shown in Table 4, the response category function evaluation of the 36-item and 32-item of WHODAS 2.0 showed that.

1. The identified average measures (−2.83 ≤ −1.58
    < −0.70 < 0.14 < 0.97 vs. −2.33 < −1.19 < −0.34
    < 0.49 < 1.42) and Andrich thresholds (τi − 2.27 <
    τi − 0.65 < τi 1.10 < τi 1.81 vs. τi − 1.90 < τi − 0.52
    < τi 0.75 < τi 1.67) for the 5-category (none, mild
difficulty, moderate difficulty, severe difficulty, and
extreme/cannot do) advanced monotonically, sug-
gest no particularly noisy category.

2. The Andrich thresholds difference of category 3
    ‘severe difficulty’ to category 4 ‘extreme/cannot do’ in
    the 36-item version, and all categories in the 32-item
    version were less than 1.4 logit, indic-
ating that the category did not fit the model.

3. The outfits MnSq of the 5-category were less than
    2.0, and therefore fit the model.

4. The probability peak of the 5-category of both
    versions did not satisfy the model expectation. The
    probability of mild, moderate, and severe difficulty
    was much lower than none, and extreme/cannot do.

The category probability curves are shown in Fig. 1. The x-axis represents what is being measured (level of disability). The y-axis represents the probability of responding to any category, ranging from 0 to 1.

(4) **Reliability and Separation**

The reliability and separation of WHODAS 2.0 and participants were satisfied and supported the internal consistencies (Table 1). The person and item reliability values of the 36-item (0.92 and 0.94) and the 32-item versions (0.95 and 0.98) were considered consistent internally. These indicate that the items measured the statistical level of disability as required. The separation index of the 36-item version (person’s separation index of 3.44; item’s separation index of 3.87) and the 32-item version (person’s separation index of 4.28; item’s separation index of 7.06) satisfied the requirement for discrimi-

ating the person and item based on their level of disability using WHODAS 2.0.

(5) **Appropriateness of the item difficulty level for the sample**

Table 5 shows the targeting statistics for the 36-item and 32-item versions of WHODAS 2.0. For the 36-item version, the average logit for persons (M = −1.84, SD = 1.54) was lower than that for items (M = 0.00, SD = 0.78) with a difference of 1.84 logit, indicating mistargeting. The targeting area between item difficulty and a person’s level of disability for this sample was 41 participants (64.1%). Twenty-five participants (35.9%) scored below the range of item difficulty (the actual level of disability was lower than that measured by WHODAS 2.0), and no participants scored above the range.

For the 32-item version, the average logit for per-
sons (M = −0.96, SD = 1.43) was lower than that for the items (M = 0.00, SD = 0.68) with a difference of 0.96 logit, indicating well-targeted. The targeting between item difficulty and person’s level of disability for this sample was 149 participants (70.3%). Fifty-six participants (26.4%) scored below the range of item difficulty (the actual level of disability was lower than that mea-
Rasch Analysis of the Indonesian version of WHODAS 2.0

sured by WHODAS 2.0), and only 7 participants (3.3%) scored above the range (the actual level of disability was higher than that measured by WHODAS 2.0). Fig. 2 shows the Wright maps to illustrate the targeting of WHODAS 2.0.

(6) **DIF of The Misfitting Item**

The DIF contrast > 0.5 of misfitting items D2.5, D3.4, and D4.5 was evaluated based on age, sex, marital status, occupational status, and assistive device usage status. In the 36-item version, DIF was found among the assistive device usage status groups for the misfitting item D2.5 (DIF contrast = –0.66, p = 0.002), and in the 32-item version, DIF was found among the marital status groups for the misfitting item D4.5 ‘sexual activity’ (DIF contrast: 0.99, p = 0.011). We did not find the DIF of D3.4 based on those characteristics’ groups.

(7) **Convergent Validity**

WHODAS 2.0 and its domains had a significantly negative correlation with WHOQOL BREF. The 36-item and 32-item version scores were strongly correlated with WHOQOL BREF scores at r = –0.65 and r = –0.64, respectively. These correlations supported the convergent validity of WHODAS 2.0.

### Table 1: Rasch model analysis of the WHODAS 2.0 construct validity

| No | WHODAS 2.0 | Sample (N) | Average outfit MnSq (SD) | Misfitting item | Variance explained by measure | Measurement reliability (Cronbach’s α) | Person-reliability | Item-reliability | Person-separation | Item-separation | WHOQOL-BREF correlation |
|----|------------|------------|-------------------------|-----------------|-------------------------------|----------------------------------------|-------------------|-----------------|-----------------|-----------------|-------------------------|
| 1  | Cognitive (domain 1) | 212 | 0.97 (0.24) | None | 69.0% | 0.94 | 0.88 | 0.96 | 2.72 | 4.92 | −0.47 |
| 2  | Getting around (domain 2) | 212 | 0.94 (0.18) | None | 79.1% | 0.93 | 0.92 | 0.99 | 3.5 | 9.04 | −0.52 |
| 3  | Self-care (domain 3) | 212 | 0.96 (0.47) | 1 item (D3.4) | 75.3% | 0.93 | 0.88 | 0.97 | 2.70 | 5.4 | −0.51 |
| 4  | Getting along with people (domain 4) | 212 | 1.07 (0.84) | 1 item (D4.5) | 60.9% | 0.83 | 0.82 | 0.83 | 2.16 | 2.37 | −0.47 |
| 5  | Life activities: household activities (domain 5) | 212 | 0.95 (0.27) | None | 79.1% | 0.94 | 0.92 | 0.98 | 3.47 | 7.87 | −0.59 |
| 6  | Life activities: work and school (domain 5) | 66 | 0.93 (0.14) | None | 81.5% | 0.93 | 0.91 | 0.97 | 3.15 | 6.05 | −0.48 |
| 7  | Participation in society (domain 6) | 212 | 0.98 (0.16) | None | 49.7% | 0.86 | 0.82 | 0.94 | 2.11 | 3.92 | −0.61 |
| 8  | 36-item version | 66 | 0.99 (0.28) | 2 items (D2.5 and D4.5) | 49.0% | 0.96 | 0.92 | 0.94 | 3.44 | 3.87 | −0.65 |
| 9  | 32-item version (four work/school items omitted) | 212 | 0.99 (0.24) | 2 items (D2.5 and D4.5) | 56.3% | 0.97 | 0.95 | 0.98 | 4.28 | 7.06 | −0.64 |

SD: Standard deviation; Cronbach’s alpha; WHODAS 2.0: the World Health Organization Disability Assessment Schedule 2.0; WHOQOL-BREF: the abbreviated version of the World Health Organization Quality of Life instrument

### Table 2: Sample characteristics

| Variables | Sample (N=212) |
|-----------|----------------|
| Age (mean ± SD) | 41.1 ± 12.2 |
| **Sex (frequency, %)** | |
| Male | 123 (58.0%) |
| Female | 89 (42.0%) |
| **Marital status (frequency, %)** | |
| Single (not married or widow or widower) | 112 (52.8%) |
| Married | 100 (47.2%) |
| **Education status (frequency, %)** | |
| Uneducated | 30 (14.2%) |
| Graduated from elementary school | 79 (37.3%) |
| Graduated from high school or above | 93 (48.6%) |
| **Occupation status (frequency, %)** | |
| Non-worker | 146 (68.9%) |
| Worker | 66 (31.1%) |
| **Type of physical Impairment (frequency, %)** | |
| Hemiplegia/paresis | 108 (50.9%) |
| Paraplegia/paresis | 40 (18.9%) |
| Tetraplegia/paresis | 11 (5.2%) |
| Other | 53 (25.0%) |
| **Assistive device usage status (frequency, %)** | |
| None | 101 (48.3%) |
| User | 108 (51.7%) |

SD: standard deviation
Table 3  Mean Square (MnSq) of the 36-item version of WHODAS 2.0

| Code | Questions                                                                 | N=212 | Infit MnSq | Outfit MnSq |
|------|---------------------------------------------------------------------------|-------|------------|-------------|
|      |                                                                           |       |            |             |
|      | **Domain 1 (cognitive):**                                                 |       |            |             |
| D1.1 | Concentrating on doing something for ten minutes?                         | 1.10  | 1.07       |             |
| D1.2 | Remembering to do important things?                                       | 0.87  | 0.86       |             |
| D1.3 | Analysing and finding solutions to problems in day-to-day life?           | 1.02  | 0.97       |             |
| D1.4 | Learning a new task, for example, learning how to get to a new place?     | 0.93  | 0.84       |             |
| D1.5 | Generally, understanding what people say?                                 | 0.72  | 0.59       |             |
| D1.6 | Starting and maintaining a conversation?                                  | 1.25  | 1.11       |             |
|      | **Domain 2 (getting around):**                                            |       |            |             |
| D2.1 | Standing for long periods such as 30 minutes?                             | 1.35  | 1.38       |             |
| D2.2 | Standing up from sitting down?                                            | 1.21  | 1.28       |             |
| D2.3 | Moving around inside your home?                                           | 0.72  | 0.72       |             |
| D2.4 | Getting out of your home?                                                 | 1.15  | 1.14       |             |
| D2.5 | **Walking a long distance such as a kilometre [or equivalent]**           | 1.88  | 1.85       |             |
|      |                                                                           | 1.46* | 1.34*      |             |
|      | **Domain 3 (self-care):**                                                 |       |            |             |
| D3.1 | Washing your whole body?                                                  | 0.53  | 0.52       |             |
| D3.2 | Getting dressed?                                                          | 0.80  | 0.79       |             |
| D3.3 | Eating?                                                                   | 0.71  | 0.68       |             |
| D3.4 | **Staying by yourself for a few days?**                                   | 0.94  | 0.83       |             |
|      |                                                                           | 1.59* | 1.76*      |             |
|      | **Domain 4 (getting along with people):**                                 |       |            |             |
| D4.1 | Dealing with people you do not know?                                      | 0.80  | 0.74       |             |
| D4.2 | Maintaining a friendship?                                                 | 0.76  | 0.69       |             |
| D4.3 | Getting along with people who are close to you?                          | 0.96  | 0.89       |             |
| D4.4 | Making new friends?                                                       | 0.86  | 0.86       |             |
| D4.5 | **Sexual activities?**                                                    | 1.80  | 1.44       |             |
|      |                                                                           | 1.94* | 1.57*      |             |
|      |                                                                           | 2.99* | 2.75*      |             |
|      | **Domain 5 (household activities)**                                       |       |            |             |
| D5.1 | Taking care of your household responsibilities?                           | 0.75  | 0.79       |             |
| D5.2 | Doing the most important household tasks well?                            | 0.58  | 0.62       |             |
| D5.3 | Getting all the household work done that you needed to do?               | 0.64  | 0.65       |             |
| D5.4 | Getting your household work done as quickly as needed?                    | 0.95  | 0.97       |             |
|      | **Domain 5 (work and school)**                                            |       |            |             |
| D5.5 | Your day-to-day work/school?                                              | 1.13  | 1.26       |             |
| D5.6 | Doing your most important work/school tasks well?                         | 0.98  | 1.08       |             |
| D5.7 | Getting all the work done that you need to do?                            | 1.19  | 1.25       |             |
| D5.8 | Getting your work done as quickly as needed?                              | 0.96  | 0.96       |             |
|      | **Domain 6 (participation in society):**                                  |       |            |             |
| D6.1 | How much of a problem did you have in joining in community activities (for example, festivities, religious or other activities) in the same way as anyone else can? | 1.03  | 0.99       |             |
| D6.2 | How much of a problem did you have because of barriers or hindrances in the world around you? | 1.08  | 1.39       |             |
| D6.3 | How much of a problem did you have living with dignity because of the attitudes and actions of others? | 0.95  | 1.00       |             |
| D6.4 | How much time did you spend on your health condition or its consequences? | 0.96  | 0.99       |             |
| D6.5 | How much have you been emotionally affected by your health condition?     | 0.94  | 1.17       |             |
| D6.6 | How much has your health been a drain on the financial resources of you or your family? | 1.06  | 1.16       |             |
| D6.7 | How much of a problem did your family have because of your health problems? | 1.17  | 1.12       |             |
| D6.8 | How much of a problem did you have in doing things by yourself for relaxation or pleasure? | 0.96  | 0.91       |             |

MnSq: Mean square; A value in bold indicates a misfitting item; *Infit and outfit MnSq logit of items (D2.5 and D4.5) that were obtained from the analysis of the 32-item version; †Infit and outfit MnSq logit of items (D3.4 and D4.5) that were obtained from a separate analysis of their domain
Table 4  Statistics of the WHODAS 2.0 response category function

| Category | 36-item version | 32-item version |
|----------|----------------|----------------|
|          | Observe average | Andrich threshold | Andrich threshold difference | Outfit MnSq | Observe average | Andrich threshold | Andrich threshold difference | Outfit MnSq |
| 0 (None) | −2.83           | NA              | NA                          | 1.03       | −2.33          | None             | NA                          | 1.13       |
| 1 (Mild) | −1.58           | −2.27           | 1.62                        | 0.79       | −1.19          | −1.90            | 1.38                        | 0.90       |
| 2 (Moderate) | −0.7        | −0.65           | 1.75                        | 1.09       | −0.34          | −0.52            | 1.27                        | 0.97       |
| 3 (Severe) | 0.14           | 1.10            | 0.71                        | 1.07       | 0.49           | 0.75             | 0.92                        | 0.95       |
| 4 (Extreme/ cannot do) | 0.97       | 1.81            | NA                          | 0.96       | 1.42           | 1.67             | NA                          | 0.90       |

NA: not available; A value in **bold**: the Andrich threshold difference < 1.4 logit

Table 5  Statistics of the WHODAS 2.0 targeting

| Data set | WHODAS 2.0 | Sample (freq) | Person logit mean (SD) | Person-Item mean difference | Targeting area (%) | Score above the range (%) | Score below the range (%) |
|----------|------------|---------------|------------------------|-----------------------------|--------------------|---------------------------|---------------------------|
| 1        | 36 items   | 83            | −1.54 (1.27)           | 1.54                        | 42.8               | 0                         | 51.8                      |
|          | 32 items   | 271           | −0.77 (1.22)           | 0.77                        | 63.5               | 4.8                       | 31.7                      |
| 2        | 36 items   | 66            | −1.84 (1.54)           | 1.84                        | 64.1               | 0                         | 35.9                      |
|          | 32 items   | 212           | −0.96 (1.43)           | 0.96                        | 70.4               | 3.2                       | 26.4                      |

Freq: frequency

Ordinal to Interval Scale Conversion

The 32-item version of WHODAS 2.0 satisfied the Rasch model expectations, supported the construct validity, and, therefore, a converted interval logit measure can be used. Table 6 presents the conversion of the raw score (0 to 128) to the Rasch measure (−6.82 to 6.53 logit).

Discussion

This is the first study to evaluate the construct validity of the self-administered Indonesian version of WHODAS 2.0 among persons with physical impairments using a Rasch model analysis. WHODAS 2.0 has been used by the Indonesian health and social ministry since 2011 to evaluate a person’s level of disability to prioritise health and welfare-oriented programs. However, the disability-related data in Indonesia are inconsistent, which challenges the validity of the WHODAS 2.0.

According to 2019 impairment data from the Social and Welfare Office, Bandung has the greatest percentage of persons with physical impairments (and combination or other impairment types) at approximately 44%. More than half of them suffered from only physical impairment (59%). Thus, we recruited persons with physical impairments to participate in this study.

We used 271 participants to evaluate the instrument. It was identified that both versions of WHODAS 2.0 were reliable, and the 32-item version was unidimensional. There were misfitting items: four items in the 36-item version (D2.1, D2.2, D2.5 and D6.7) and four items in the 32-item version (D4.5, D5.4, D6.1 and D6.7) that might reduce their overall performance. According to guidance, participants with missing values > 30% and participants with infit or outfit MnSq > 1.4 (misfitting person) were necessarily excluded to prevent bias in data analysis. Misfitting persons are those who have unexpected responses to the questions. It could be due to special knowledge, lucky guessing, carelessness, mistakes, misunderstanding or exceptional ignorance.

The results provided evidence for several aspects of validity. After the 59 data were excluded, the principal components analysis (PCA) analysis using the 212 remaining participants (66 working participants) showed that the instrument was unidimensional. The 36-item version was not considered unidimensional because the variance explained by the measure was 49%. However, it was very close to the desired 50% of the variance explained by the measure. The explanation could be that this study’s working participant sample size was not adequate for the data analysis. Further analysis using a larger sample might be necessary to evaluate the 36-item version. The 32-item version fit the unidimensionality of the Rasch model as the variance explained by the measure reached 56.3%.

The average MnSq 0.99 logit of WHODAS 2.0 fit the model. However, two misfitting items were identified from both versions: D2.5 ‘walking a long distance such as a kilometre [or equivalent]’ and D4.5 ‘sexual activity’. Item D3.4 ‘Staying by yourself for a few days’ was found as a misfitting item when separately analysed in domain 3, ‘self-care’. These results are consistent with previous studies. The discovered misfitting items signified that the performance on these items could not be predicted comfortably by what is known about these persons’ abilities based on the overall WHODAS 2.0 total score. A possible explanation could be that these questions do not represent the same construct or are confounded by other factors.

The misfitting item D2.5 ‘walking a long distance such as a kilometre [or equivalent]’ relates to the experience of persons with disability using their legs for walking. Among misfitting persons, 78% had paresis/plegia in their legs, making it difficult for them to walk. More-
Fig. 1  Category probability curves of the WHODAS 2.0 response category
Fig. 2 Wright map of the person-item distribution. A is the Wright map of the 36-item version, and B is for the 32-item version (worker items omitted). Each "#" represents 2 subjects, each "." or "x" represents 1 subject (M = mean; S = 1 standard deviation from the mean; T = 2 standard deviations from the mean).
over, 57% of them did not have any assistive devices supporting them in daily activities. They might have no experience of undertaking such activities as walking a long distance. Therefore, it is not surprising that responses to this item diverged from the expected difficulty pattern based on the overall performance of WHODAS 2.0. According to the assistive device status group, the DIF evaluation of item D2.5 also showed bias and might function differently.

The finding of D3.4 ‘ ‘Staying by yourself for a few days’ as a misfitting item is consistent with the result obtained by De Wolf et al. The participants need assistance for doing daily activities such as eating, bathing, dressing, and bowel and bladder care. This finding explains why they could not stay home alone for more than one day, despite experiencing low levels of disability in everyday life. The misfitting item D4.5 ‘ ‘sexual activities’ is also similar to the result obtained by De Wolf et al. It could be related to the missing response to the item as a result of the person’s discouragement in response to the item. 21.2% of participants did not answer question D4.5. Unmarried (30%) and widows or widowers (30%) participants were among those who did not respond to the question (47.4%). On the other hand, only 6% of married participants did not respond to this question. The DIF evaluation on item D4.5 showed bias on marital status. The single participants had difficulty answering item D4.5 compared to those who were married. These conditions might be related to Indonesian social norms. In Indonesia, premarital sex is socially inappropriate and taboo. Thus, choosing not to answer the sexual-related questions might be the best way for them, especially for singles (unmarried and widow/widower).

WHODAS 2.0 revealed good to excellent internal consistency for our sample, with a Cronbach’s alpha ranging from 0.85 to 0.97. This finding is similar to the previous studies conducted in Indonesia, Taiwan, and the United States. The separation index of both versions of WHODAS 2.0 satisfied the requirement for discriminating the persons and items based on their disability level. The response category scale of WHODAS 2.0 did not perfectly fit the Rasch model. This result aligns with the study results of Küçükdeveci et al. who reported unordered thresholds and collapsed the response categories. In this study, we did not collapse the categories because it still adequately performed for unidimensionality, differentiated person’s levels of disability across the continuum of disability, and precisely distinguished the item and person’s levels of disability in our physical impairment sample. However, collapsing or combining the specific response categories would be necessary to optimise the measurement effectiveness in the future, especially for the 36-item version.

Even though the 36-item and 32-item versions’ targeting reached 64.1% and 70.3%, respectively, some items showed a mismatch in the range of a person’s disability level, especially those with less disability. These findings are in accordance with Herrold et al., in which some items are too easy and are not well-targeted to the sample. Further research involving persons with higher levels of disability is necessary.

Consistent with the study by Ustun et al., convergent validity for the WHODAS 2.0 was confirmed with the expected correlation with WHOQOL BREF. Strong negative correlation coefficients were observed between WHODAS 2.0 and WHOQOL BREF scores. These findings indicate that the Indonesian version of WHODAS 2.0 properly measured the level of disability in our sample.

The ordinal scale of the 36-item version of WHODAS 2.0 was not converted to a logit scale because it was not unidimensional. The 32-item version was unidimensional and, therefore the Rasch model analysis converted the ordinal summed scores to interval logit measures. We provide Table 6 to show the conversion scores of the 32-item version. Thus, even if the WHODAS 2.0 users are not familiar with Rasch model analysis, they can directly use the logit measures we provided in the table.

We acknowledge that this study has the following limitations. The participant’s level of disability was not equally distributed to enough participants with a higher level of disability. The sample size of workers in this study was insufficient to evaluate the 36-item version and domain 5 (work and school). WHODAS 2.0 is used for assessing disability levels in persons with various impairments, such as physical, vision, cognitive, and hearing impairments. However, because this study focused on persons with physical impairments, the scope of invariant measurement analysis may be limited. We could not analyse the equality of the relationship between the level of disability (latent trait) and the observed scores across the persons with various impairments. Therefore, whether they were different at certain disability levels could not be determined. We recruited participants with physical impairments from one Indonesian region (Bandung). This could limit the generalizability of the findings to other regions of Indonesia. However, we believe that these findings will contribute to expanding impairment and disability-related evidence in Indonesia.

We conclude that even though the Indonesian version of WHODAS 2.0 has misfitting items, an unsatisfactory response category function, and is not fully well-targeted to the sample, Rasch model analysis reveals that the instrument is unidimensional and has satisfactory internal consistency. These results indicate that the Indonesian version of WHODAS 2.0, mainly the 32-item version, has acceptable construct validity to assess the level of disability in a physical impairment sample in Bandung, Indonesia. Further analysis would be necessary to guarantee the construct validity assessment results of the 36-item version of WHODAS 2.0.

**Acknowledgements**

The authors appreciate the persons with physical impairment for their participation in the study. Special
thanks are given to the WHODAS 2.0 group to approve the Indonesian language version of WHODAS 2.0. This work was proportionally supported by the Research Activity Activation Program 2019 of the Graduate School of Medicine, Gunma University, and the Unpad Doctoral Dissertation Research of the Universitas Padjadjaran 2019 (Grant number: 7876/UN6.C/LT/2019).

Declaration of Interest: The authors declare no conflicts of interest.

Table 6 Interval score conversion of the 32-item version of WHODAS 2.0

| Ordinal scores (raw sum scores) | Rasch measures (logit) | Ordinal scores (raw sum scores) | Rasch measures (logit) |
|---------------------------------|------------------------|---------------------------------|------------------------|
| 0                               | −6.82                  | 65                              | 0.10                   |
| 1                               | −5.60                  | 66                              | 0.14                   |
| 2                               | −4.88                  | 67                              | 0.18                   |
| 3                               | −4.45                  | 68                              | 0.23                   |
| 4                               | −4.13                  | 69                              | 0.27                   |
| 5                               | −3.88                  | 70                              | 0.31                   |
| 6                               | −3.68                  | 71                              | 0.35                   |
| 7                               | −3.50                  | 72                              | 0.39                   |
| 8                               | −3.34                  | 73                              | 0.43                   |
| 9                               | −3.20                  | 74                              | 0.47                   |
| 10                              | −3.07                  | 75                              | 0.51                   |
| 11                              | −2.95                  | 76                              | 0.56                   |
| 12                              | −2.84                  | 77                              | 0.60                   |
| 13                              | −2.74                  | 78                              | 0.64                   |
| 14                              | −2.64                  | 79                              | 0.68                   |
| 15                              | −2.55                  | 80                              | 0.72                   |
| 16                              | −2.46                  | 81                              | 0.77                   |
| 17                              | −2.37                  | 82                              | 0.81                   |
| 18                              | −2.29                  | 83                              | 0.85                   |
| 19                              | −2.22                  | 84                              | 0.89                   |
| 20                              | −2.14                  | 85                              | 0.94                   |
| 21                              | −2.07                  | 86                              | 0.98                   |
| 22                              | −2.00                  | 87                              | 1.02                   |
| 23                              | −1.93                  | 88                              | 1.07                   |
| 24                              | −1.87                  | 89                              | 1.11                   |
| 25                              | −1.81                  | 90                              | 1.16                   |
| 26                              | −1.74                  | 91                              | 1.20                   |
| 27                              | −1.68                  | 92                              | 1.25                   |
| 28                              | −1.62                  | 93                              | 1.29                   |
| 29                              | −1.57                  | 94                              | 1.34                   |
| 30                              | −1.51                  | 95                              | 1.39                   |
| 31                              | −1.45                  | 96                              | 1.44                   |
| 32                              | −1.40                  | 97                              | 1.49                   |

| Ordinal scores (raw sum scores) | Rasch measures (logit) | Ordinal scores (raw sum scores) | Rasch measures (logit) |
|---------------------------------|------------------------|---------------------------------|------------------------|
| 33                              | −1.35                  | 98                              | 1.54                   |
| 34                              | −1.29                  | 99                              | 1.59                   |
| 35                              | −1.24                  | 100                             | 1.64                   |
| 36                              | −1.19                  | 101                             | 1.69                   |
| 37                              | −1.14                  | 102                             | 1.75                   |
| 38                              | −1.09                  | 103                             | 1.80                   |
| 39                              | −1.04                  | 104                             | 1.86                   |
| 40                              | −0.99                  | 105                             | 1.92                   |
| 41                              | −0.94                  | 106                             | 1.98                   |
| 42                              | −0.90                  | 107                             | 2.04                   |
| 43                              | −0.85                  | 108                             | 2.10                   |
| 44                              | −0.80                  | 109                             | 2.17                   |
| 45                              | −0.76                  | 110                             | 2.23                   |
| 46                              | −0.71                  | 111                             | 2.31                   |
| 47                              | −0.67                  | 112                             | 2.38                   |
| 48                              | −0.62                  | 113                             | 2.46                   |
| 49                              | −0.58                  | 114                             | 2.54                   |
| 50                              | −0.53                  | 115                             | 2.63                   |
| 51                              | −0.49                  | 116                             | 2.72                   |
| 52                              | −0.45                  | 117                             | 2.82                   |
| 53                              | −0.40                  | 118                             | 2.93                   |
| 54                              | −0.36                  | 119                             | 3.04                   |
| 55                              | −0.32                  | 120                             | 3.17                   |
| 56                              | −0.27                  | 121                             | 3.32                   |
| 57                              | −0.23                  | 122                             | 3.48                   |
| 58                              | −0.19                  | 123                             | 3.68                   |
| 59                              | −0.15                  | 124                             | 3.91                   |
| 60                              | −0.11                  | 125                             | 4.21                   |
| 61                              | −0.06                  | 126                             | 4.62                   |
| 62                              | −0.02                  | 127                             | 5.32                   |
| 63                              | 0.02                   | 128                             | 6.53                   |
| 64                              | 0.06                   |                                  |                        |

Logit: log-odds unit

References

1. World Health Organization (ed). The International Classification of Functioning, Disability and Health. Vol. 18. Geneva: World Health Organization (WHO), 2006.
2. World Health Organization and World Bank (eds). World Report on Disability. Vol. 14. Geneva: World Health Organization (WHO), 2011.
3. Corporation for National and Community Service (ed). Creating an Inclusive Environment: A Handbook for the Inclusion of People with Disabilities in National and Community Service Programs. California: The UCP Access AmeriCorps, 2004.
4. World Health Organization (ed). Measuring Health and Dis-
ability: Manual for WHO Disability Assessment Schedule WHODAS 2.0. Geneva: World Health Organization (WHO), 2010.

5. The Sustainable Development Solutions Network Leadership Council (ed). Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a Data Revolution. New York: the SDSN Leadership Council, 2015.

6. The United Nations (ed). Convention on the Rights of Persons with Disabilities Advocacy Toolkit. 15th ed. Geneva: the United Nations, 2008.

7. The United Nations (ed). Convention on the Rights of Persons with Disabilities and Optional Protocol. Vol. 1. Geneva: the United Nations, 2006.

8. World Health Organization (ed). How to Use the ICF: A Practical Manual for Using the International Classification of Functioning, Disability and Health (ICF). Geneva: World Health Organization (WHO), 2013.

9. Adioetomo S.M, Mont D, Irwanto. Persons with Disabilities in Indonesia: Empirical Facts and Implications for Social Protection Policies. Jakarta: TNP2K, 2014.

10. Wright BD, Mok MM. An Overview of the Family of Rasch Measurement Models. In: Smith Jr E V., Smith RM. (eds). Introduction to Rasch Measurement: Theory, Models, and Applications. Maple Grove: JAM Press, 2004: 1-24.

11. Baghaei P. The Rasch model as a construct validation tool. Rasch Meas Trans 2008; 22: 1145-1162.

12. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Heal Pharm 2008; 65: 2276-2284.

13. Taherdoost H. Validity and reliability of the research instrument: How to test the validation of a questionnaire/survey in a research. SSRN Electron J 2016; 5: 28-36.

14. Pitts RT, Naumenko O. The 2014 standards for educational and psychological testing: What teachers initially need to know. Work Pap Educ 2016: 2: 1-6.

15. Magni C. Demonstrating the difference between classical test theory and item response theory using derived test data. Int J Educ Psychol Assess 2009; 1: 1-11.

16. Waugh RF, Chapman ES. An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: what is the difference? Which method is better? J Appl Meas 2005; 6: 80-99.

17. Impara JC, Plake BS. Teachers’ ability to estimate item difficulty: A test of the assumptions in the Angoff standard setting method. J Educ Meas 1998; 35: 69-81.

18. Pösö M, Cieza A, Stucki G. Psychometric properties of the WHODAS II in rehabilitation patients. Qual Life Res 2007; 16: 1521-1531.

19. Schlote A, Richter M, Wunderlich MT, et al. Der WHODAS II in der anwendung bei schlaganfallpatienten und ihren angehörigen: Rehabilität und inter-rater-reliabilität. Rehabilitation 2008; 47: 31-38.

20. Chwastiak LA, Von Korff M. Disability in depression and back pain: Evaluation of the World Health Organization Disability Assessment Schedule (WHODAS II) in a primary care setting. J Clin Epidemiol 2003; 56: 507-514.

21. Yuliana S, Mushli M, Sim J, et al. Development and validation of the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) Indonesian version in stroke survivors. Disabil Rehabil 2021: 1-8.

22. Santor D, Ramsay J. Progress in the technology of measurement: Applications of item response models. Psychol Assess 1998; 10: 345-359.

23. Prieto L, Alonso J, Lamarca R. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. Health Qual Life Outcomes 2003; 1: 1-13.

24. Rusch T, Lowry PB, Mair P, et al. Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. Inf Manag 2017; 54: 189-203.

25. Boone WJ, Yale MS, Stuyver JR. Rasch Analysis in the Human Sciences. Oxford: Springer, 2014.

26. Hambleton RK, Van Der Linden WJ. Advances in item response theory and applications: An introduction. Appl Psychol Meas 1982: 6: 373-378.

27. Mitchell-Parker K, Medvedev ON, Krägeloh CU, et al. Rasch analysis of the Frost Multidimensional Perfectionism Scale. Aust J Psychol 2018; 70: 258-268.

28. Wright BD, Masters GN. Rating Scale Analysis: Rasch Measurement. Chicago: MESA PRESS, 1982.

29. Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. Value Heal 2004; 7: 22-26.

30. Üstün TB, Chatterji S, Kontajsek N, et al. Developing the World Health Organization Disability Assessment Schedule 2.0. Bull World Health Organ 2010; 88: 815-823.

31. De Wolf AC, Tate RL, Lannin NA, 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: 747-755.

32. Snell DL, Siegert RJ, Silverberg ND. Rasch analysis of the World Health Organization Disability Assessment Schedule 2.0 in a mild traumatic brain injury sample. Brain Inj 2020; 34: 610-618.

33. Chiu TY, Finger ME, Fellinghauer CS, et al. Validation of the World Health Organization Disability Assessment Schedule 2.0 in adults with spinal cord injury in Taiwan: A psychometric study. Spinal Cord 2019; 57: 516-524.

34. Kutlay Ş, Küküçüdevci AA, Elhan AH, et al. Validation of the World Health Organization disability assessment schedule II (WHODAS-II) in patients with osteoarthritis. Rheumatol Int 2011; 31: 339-346.

35. Tarigan AKM, Sagala S, Samsura DAA, et al. Bandung City, Indonesia. Cities 2016; 50: 100-110.

36. Indonesian Ministry of Health (ed). Laporan Nasional Riske das 2018 [Risksedas National Report 2018]. Jakarta: Indonesian Ministry of Health, 2018.

37. Anjara SG, Bonetto C, Ganguli P, et al. Can general practitioners manage mental disorders in primary care? A partially randomised, pragmatic, cluster trial. PLoS One 2019; 14: 1-26.

38. Isfandari S, Pangaribuan L, Idaiani S. Non communicable disease (NCD) as risk for disability: Recommendation for Indonesian UHIC program. Glob J Health Sci 2020; 12: 118.

39. Pujilestari CU, Nystroëm L, Norberg M, et al. Association between changes in waist circumferences and disability among older adults: WHO-INDEPTH study on global ageing and adult health (SAGE) in Indonesia. Obes Res Clin Pract 2019; 1: 1-6.

40. World Health Organization (ed). The World Health Organization Quality of Life (WHOQOL)-Bref. Geneva: World Health Organization (WHO), 2004.

41. Ch Salim O, Sudharmar NA, Kusumaratna RK, et al. Validity and reliability of World Health Organization Quality of Life-BREF to assess the quality of life in the elderly. Universa Medica 2007; 26: 27-38.

42. Purba FD, Hunfeld IAM, Iskandarsyah A, et al. Quality of life of the Indonesian general population: Test-retest reliability and population norms of the EQ-5D-5L and WHOQOL-BREF. PLoS One 2018; 13: 1-20.

43. Messick S. Test validity: A matter of consequence. Soc Indic Res 1998; 45: 35-44.

44. Fernández-Ballesteros R, Eignor DR. Standard for Educational and Psychological Testing. 2nd ed. Washington DC: American Educational Research Association, 2004.
45. Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 3rd ed. New York: Routledge, 2015.
46. Wright B, Linacre JM. Reasonable mean-square fit values. Rasch Meas Trans 1994; 8: 370.
47. Dodeen H. The relationship between item parameters and item fit. J Educ Meas 2004; 41: 261-270.
48. Linacre JM. A User’s Guide to WINSTEPS® MINISTEP Rasch-Model Computer Programs Program Manual 3.75. Chicago: Winsteps®, 2012.
49. Herrold AA, Kletzel SL, Mallinson T, et al. Psychometric measurement properties of the World Health Organization Disability Assessment Schedule 2.0 (WHODAS) evaluated among veterans with mild traumatic brain injury and behavioral health conditions. Disabil Rehabil 2019; 43: 1-10.
50. Van Der Velde G, Beaton D, Hogg-Johnston S, et al. Rasch analysis provides new insights into the measurement properties of the neck disability index. Arthritis Care Res 2009; 61: 544-551.
51. Baghaei P, Amrahi N. Validation of a multiple choice English vocabulary test with the Rasch model. J Lang Teach Res 2011; 2: 1052-1060.
52. Linacre J. Understanding Rasch measurement: Optimizing rating scale category effectiveness. J Appl Meas 2002; 3: 85-106.
53. Fink A. Survey research methods. Appl Soc Res Methods 2010; 1: 152-160.
54. Stelmack J, Szyk JP, Stelmack T, et al. Use of Rasch person-item map in exploratory data analysis: A clinical perspective. J Rehabil Res Dev 2004; 41: 233-241.
55. Van Zile-Tamsen C. Using Rasch analysis to inform rating scale development. Res High Educ 2017; 58: 922-933.
56. Lim SM, Rodger S, Brown T. Using Rasch analysis to establish the construct validity of rehabilitation assessment tools. Int J Ther Rehabil 2009; 16: 251-260.
57. Pesudovs K, Burr JM, Harley C, et al. The development, assessment, and selection of questionnaires. Optom Vis Sci 2007; 84: 663-674.
58. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. Med Care 2004; 42: 17-24.
59. Arias González VB, Crespo Sierra MT, Arias Martínez B, et al. An in-depth psychometric analysis of the Connor-Davidson Resilience Scale: Calibration with Rasch-Andrich model. Health Qual Life Outcomes 2015; 13: 1-13.
60. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998; 28: 551-558.
61. The WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL): Position paper from the World Health Organization. Soc Sci Med 1995; 41: 1403-1409.
62. Power MJ, Green AM. Development of the WHOQOL disabilities module. Qual Life Res 2010; 19: 571-584.
63. Perline R, Wright BD, Wainer H. The Rasch model as additive conjoint measurement. Appl Psychol Meas 1979; 3: 237-255.
64. Batche CS, Tennant A, Thomnard JL. ACTIVLIM-stroke: A cross-cultural Rasch-built scale of activity limitations in patients with stroke. Stroke 2012; 43: 815-823.
65. Rajeziesfahani S, Federici S, Bacci S, et al. Validity of the 36-item Persian (Farsi) version of the World Health Organization Disability Assessment Scale (WHODAS) 2.0. Int J Ment Health 2019; 48: 14-39.
66. Kucukdeveci AA, Kutlay S, Yildizlar D, et al. The reliability and validity of the World Health Organization Disability Assessment Schedule (WHODAS-II) in stroke. Disabil Rehabil 2013; 35: 214-220.
67. Pallant JF, Tennant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 2007; 46: 1-18.