



# Simplified Post-stroke Functioning Assessment Based on ICF via Dichotomous Mokken Scale Analysis and Rasch Modeling

Chun Feng<sup>1</sup>, Zhong-Li Jiang<sup>2</sup>, Ming-Xue Sun<sup>3</sup> and Feng Lin<sup>2\*</sup>

<sup>1</sup> The First Rehabilitation Hospital of Shanghai, Rehabilitation Hospital Affiliated to Tongji University, Shanghai, China, <sup>2</sup> Sir Run Run Hospital, Nanjing Medical University, Nanjing, China, <sup>3</sup> Maternal and Child Health Care Hospital of Jiangyin, Jiangyin, China

## OPEN ACCESS

### Edited by:

Margit Alt Murphy,  
University of Gothenburg, Sweden

### Reviewed by:

Guna Bērziņa,  
Rīga Stradiņš University, Latvia  
Lena Rafsten,  
University of Gothenburg, Sweden

### \*Correspondence:

Feng Lin  
peterduus@njmu.edu.cn

### Specialty section:

This article was submitted to  
Stroke,  
a section of the journal  
Frontiers in Neurology

Received: 01 December 2021

Accepted: 15 March 2022

Published: 14 April 2022

### Citation:

Feng C, Jiang Z-L, Sun M-X and Lin F  
(2022) Simplified Post-stroke  
Functioning Assessment Based on  
ICF via Dichotomous Mokken Scale  
Analysis and Rasch Modeling.  
Front. Neurol. 13:827247.  
doi: 10.3389/fneur.2022.827247

**Purpose:** This study aims to accomplish two tasks for *International Classification of Functioning, Disability and Health* (ICF) application among persons with stroke: (1) to make an ICF tool for measuring personal abilities with simplified assessment operations; (2) to quantitatively evaluate ICF categories for being functioning rather than being disabled.

**Methods:** A total of 130 inpatients with stroke *via* convenience sampling were evaluated by the extended comprehensive ICF core set for stroke, modified Rankin scale, and modified Barthel index (MBI). This study investigated the responses to 118 stroke-related ICF items (59 items in b and d domains individually) using Mokken scale analysis followed with Rasch modeling.

**Results:** A Mokken scale with 47 items was extracted from the binary data (1 as no-impairment or mild-impairment and 0 as moderate to complete impairment). A Rasch model with 45 items was derived from the Mokken scale. The conversion chart was available involving the original ordinal scores to Rasch-transformed scores from 0 to 100 (interval scale). Total scores exhibited a high correlation with the personal abilities estimated by the Rasch model. The personal ability also demonstrated a significantly strong correlation with the score of the MBI. Thus, the 45 ICF items were suggested to rate potential functional ability as a single measurement.

**Conclusion:** Based on simple “functioning or disabled” judgment tasks, ICF assessment can be simplified to a questionnaire with answering “yes-or-no” questions for each category. Functioning level for each person and difficulty of being functioning for each category can be estimated by the Rasch model of this questionnaire.

**Keywords:** Mokken scale analysis, Rasch modeling, item response theory, International Classification of Functioning, Disability and Health (ICF), ICF core set, stroke

## INTRODUCTION

At present, available functional assessments in the clinical practice include the Barthel Index (BI) for basic activities of daily-living evaluation without considering cognition status, which also contains a floor and ceiling effect for assessing activity and participation (1). On the other hand, the 36-Item Short-Form Health Survey (SF-36) involving 8 scales is the most extensively quality-of-life assessment. However, its developers point out that the total score of SF-36 cannot be a single measure for quality of life (2). The widely used scale such as the National Institutes of Health stroke scale was originally designed for acute care settings but might not be appropriate for rehabilitation practices (3).

Moreover, at present, certain assessment scales are mostly developed for specific diseases. It causes difficulty to obtain assessment with a universal standard for patients with multiple diseases (such as metabolic syndrome). Patients with stroke often suffer from comorbidities such as hypertension or diabetes (4). Patients might also experience a variety of functional disorders during the continuum of the diseases. Even patients with good recovery as assessed by traditional ways can still retain cognitive, emotional, or social integration impairments (5). Therefore, a sufficiently comprehensive assessment with high efficiency is required for the status determination of patients' activity and participation.

In 2001, the *International Classification of Functioning, Disability and Health* (ICF) as a theoretical framework and classification system was promulgated by WHO. ICF is designed to describe human experience regarding health under the umbrella terms of functioning and disability (6). The ICF qualifier scale is a 5-point Likert scale with numeric rating ranks "no impairment" = 0; "mild impairment" = 1, "moderate impairment" = 2, "severe impairment" = 3, and "complete impairment" = 4. There are several controversies regarding how to carry out such assessments (7, 8). (1) Consensuses are not achieved regarding the rating standard, especially in the determination of rating 2–4. (2) The vast number of ICF items for each disease and each item requires to rate from qualifier 0–4, which is too overwhelming to apply into clinical evaluation. (3) The core sets cannot provide a single measurement for the individual's functional level simply by the sum of scores. (4) The categories are not estimated by their difficulty or easiness of being functioning or being disabled. All categories are treated as non-hierarchic items in the core sets. The former two problems are about the efficiency of assessment. The latter two limit the sufficiency of assessment. The efficiency means that the task is time-saving. The sufficiency implies that the result is comprehensive with details. The currently available ICF core sets incline to the sufficiency rather than efficiency. It is a challenge to make a simplified comprehensive assessment tool based on the comprehensive core set.

To measure personal ability and to quantitatively differentiate item difficulties, scholars introduced the item response theory (IRT) and utilized its parametric method of Rasch modeling in several ICF studies, namely, the brief ICF core set (9, 10), rehabilitation set (11), spinal cord injury core set (12, 13), and Lucerne ICF-based multidisciplinary observation scale (14).

However, these models did not solve the problem of efficiency. The 5-point qualifier system remained unchanged. In recent years, non-parametric IRT models based on Mokken scale analysis (MSA) have begun to attract attention in the medical areas (15–17). In comparison with the necessary Guttman hypothesis held by the Rasch model, which means that highly competent subjects must be bound to score on the easy tasks (18), MSA holds the probabilistic Guttman hypothesis that high-ability subjects were more likely to complete low-difficulty tasks. Several conditions, namely, single scale, local independence, monotonicity, and item invariant item ordering (IIO) can be feasibly checked by MSA. These are conditions required for further parametric IRT processes, especially Rasch modeling (19). That is, the MSA offers preparation processes of data shaping and hypothesis testing for Rasch modeling.

The purpose of this multicenter, cross-sectional study was to provide an ICF-based dichotomous-scoring scale and its relative Rasch model to assess personal ability and item difficulty among the Chinese stroke population. In comparison with the five-point scoring ICF system, the dichotomous-scoring scale was simplified. We assumed that the final scale with high reliability and validity was based on the MSA and Rasch modeling. The simplified ICF scale might be a promising tool for evaluating individual functional levels. The final Rasch model offered an estimation of difficulty being functional for each item.

## METHODS

### Subject Recruitment and Study Design

This study is a multicenter and cross-sectional study from October of 2018 to June of 2020 involving 130 inpatients' acute, subacute, or chronic phases of recovery from stroke by convenience sampling. To increase the result generalizability for convenience sampling, we followed a maximum variation sampling strategy to capture the common patterns from a great deal of heterogeneity (20).

After providing written informed consent, participants were asked to complete a one-h interview and clinical examination in a private room. The demographic information collected included gender, age, marital status, and education level. The diagnostic category, recurrence, plegia side, body mass index, and the duration of disease were reviewed through medical charts or face-to-face interviews. Based on the interview and clinical examination, two trained evaluators were administered to evaluate modified Rankin scale (MRS) and Modified Barthel Index (MBI) and rate 0–4, 8 (no specified), or 9 (not applicable) for the extended comprehensive ICF core set for stroke questionnaire (21). The ICF categories scored 8 and 9 were defined as items with missing data. Persons with missing ratios > 30% were excluded because there was no reliably unbiased missing value imputation method (22). The eligibility criteria included (21): (1) diagnosis of stroke by the computed tomography (CT) or magnetic resonance imaging (MRI); (2) stroke with plegia as the main diagnosis; (3) age  $\geq$  18 years old; and (4) able to provide the informed consent. The exclusion criteria: (1) unhealed trauma or surgical incision; (2) patients with critical illness, such as cardiopulmonary failure; and (3)

other diseases affecting data collection, such as a history of mental illness or severe dementia.

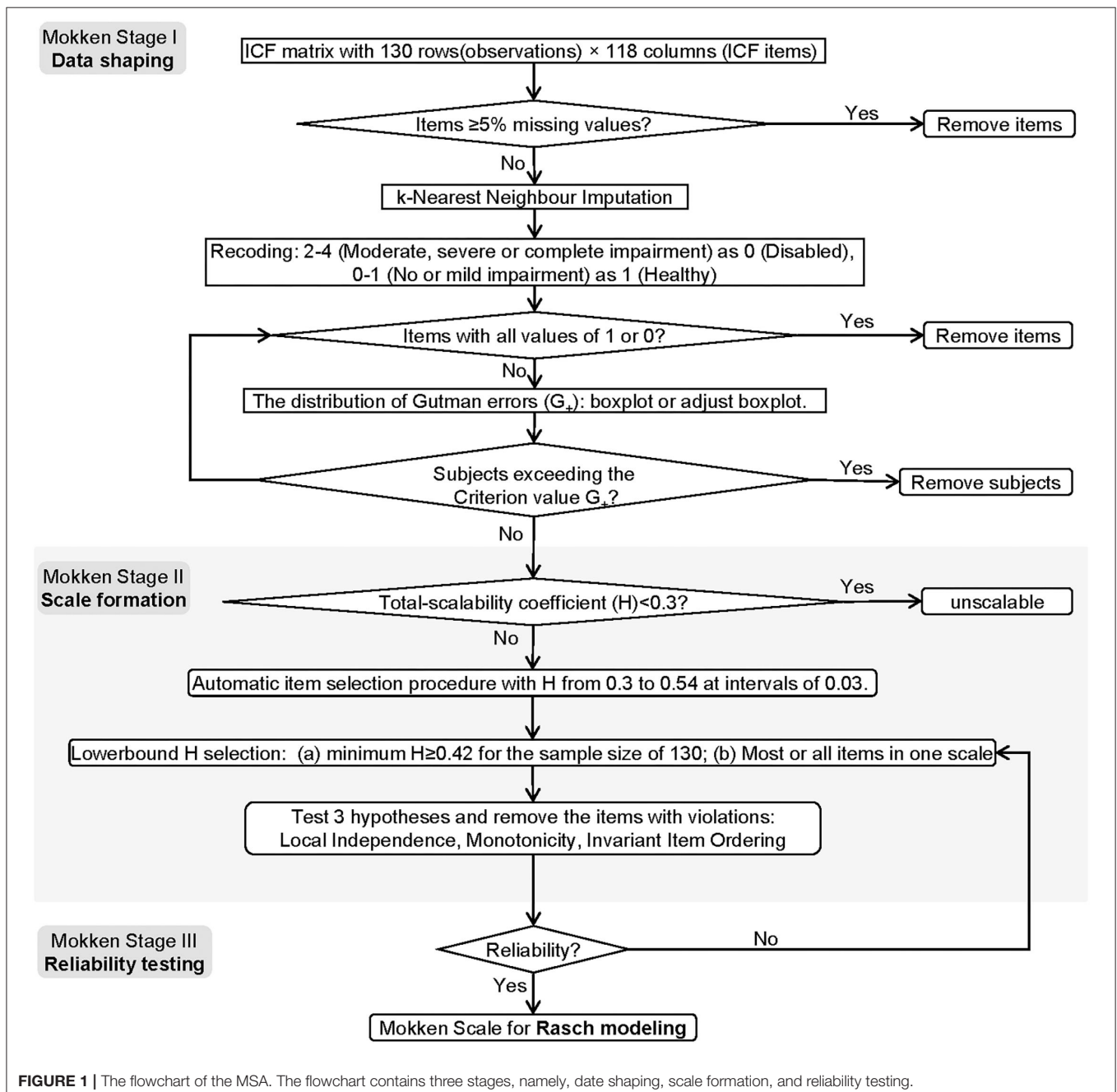
Extended comprehensive ICF core set for stroke consists of 166 items (21), namely, 59 items related to body function (b), 59 items to activity and participation (d), 37 items to environmental and individual factors (e), and 11 items to body structure (s). This study sorted out the b and d categories (**Appendix 1**) and excluded e and s categories. The main reason was that this study was completed in the inpatient setting. There was the same tendency of e categories for the patients admitted to the hospital. For instance, items regarding the information of health practitioners, relatives, and families tended to attain

similar scores in the conditions of hospitalization. In addition, the qualifier of e categories is scored differently than b, d, and s. The s categories with non-interventional items, such as the structure of the brain, were also excluded.

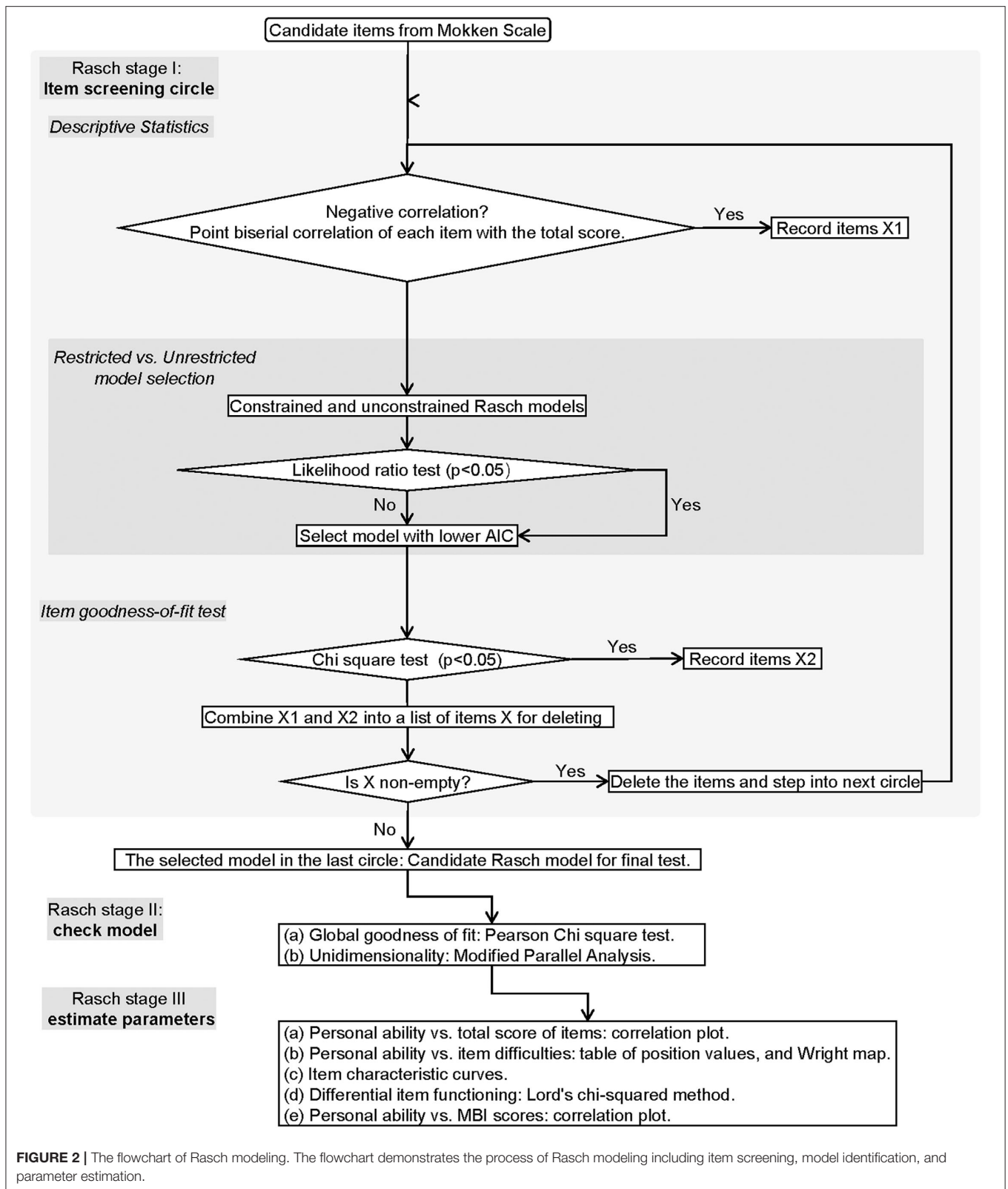
This study was approved by the ethics committee of two local rehabilitation hospitals in Shanghai and Nanjing and informed consent was obtained from all subjects before the experiment.

## Data Analysis

The brief flowchart of MSA and the Rasch modeling were exemplified in **Figures 1, 2**. The MSA utilized the “mokken” package of R and the guideline of Sijtsma and van der Ark (23).



**FIGURE 1** | The flowchart of the MSA. The flowchart contains three stages, namely, data shaping, scale formation, and reliability testing.



**FIGURE 2 |** The flowchart of Rasch modeling. The flowchart demonstrates the process of Rasch modeling including item screening, model identification, and parameter estimation.

**TABLE 1** | Description of demographic data and characteristics of the disease [mean (SD) or number (percent)].

Factor	Level	Both	Female	Male	<i>p</i>
Gender		130 (100%)	36 (27.7%)	94 (72.3%)	
Education time	<6 years	12 (9.23%)	6 (50.0%)	6 (50.0%)	0.324
	6–9 years	48 (36.9%)	13 (27.1%)	35 (72.9%)	
	10–12 years	33 (25.4%)	9 (27.3%)	24 (72.7%)	
	>12 years	37 (28.5%)	8 (21.6%)	29 (78.4%)	
Solitary	No	105 (80.8%)	28 (26.7%)	77 (73.3%)	0.774
	Yes	25 (19.2%)	8 (32.0%)	17 (68.0%)	
Diagnosis	Hemorrhage	34 (26.2%)	12 (35.3%)	22 (64.7%)	0.475
	Infarction	95 (73.1%)	24 (25.3%)	71 (74.7%)	
	Both	1 (0.77%)	0 (0.00%)	1 (100%)	
Recurrence	No	93 (71.5%)	26 (28.0%)	67 (72.0%)	1.000
	Yes	37 (28.5%)	10 (27.0%)	27 (73.0%)	
Plegia side	Left	60 (46.2%)	20 (33.3%)	13 (86.7%)	0.286
	Right	55 (42.3%)	14 (25.5%)	40 (66.7%)	
	Both	15 (11.5%)	2 (13.3%)	41 (74.5%)	
Modified rankin scale	1	11 (8.46%)	1 (9.09%)	10 (90.9%)	0.675
	2	19 (14.6%)	6 (31.6%)	13 (68.4%)	
	3	21 (16.2%)	5 (23.8%)	16 (76.2%)	
	4	65 (50.0%)	20 (30.8%)	45 (69.2%)	
	5	14 (10.8%)	4 (28.6%)	10 (71.4%)	
Post-stroke stage	≤2 weeks	17 (13.1%)	2 (5.56%)	15 (16.0%)	0.212
	3–27 weeks	94 (72.3%)	30 (83.3%)	64 (68.1%)	
	≥28 weeks	19 (14.6%)	4 (11.1%)	15 (16.0%)	
Age (years)		64.9 (13.0)	67.89 (14.00)	63.70 (12.43)	0.1213
Course of disease (months)		4.46 (8.62)	3.22 (4.12)	4.94 (9.79)	0.1629
Modified Barthel Index (scores)		47.7 (25.7)	45.97 (24.69)	48.35 (26.20)	0.6305
Body mass index (kg/m <sup>2</sup> )		24.3 (3.85)	23.92 (4.43)	24.45 (3.61)	0.5213

Gender differences were estimated by  $\chi^2$  test for categorical data and *t*-test for continuous data.

The Rasch modeling was with the “lrm” package of R (24). The study method was reviewed by an expert in the IRT field.

## MSA Stages

### MSA I: Data Shaping

To shrink imputation size, we applied a relatively conservative cut value (item missing values  $\geq 5\%$ ) to exclude the categories. The censored data were completed by *k*-nearest neighbor imputation with *k* = 5. The imputed data were then binarized using the following criteria: no or mild impairment as 1 (functioning), and moderate, severe, or complete impairment as 0 (disabled). We excluded the participants who were outliers in the distribution of the number of Guttman’s errors ( $G_+$ ) (22). Constantly valued items should be removed for controlling ceiling or floor effects.

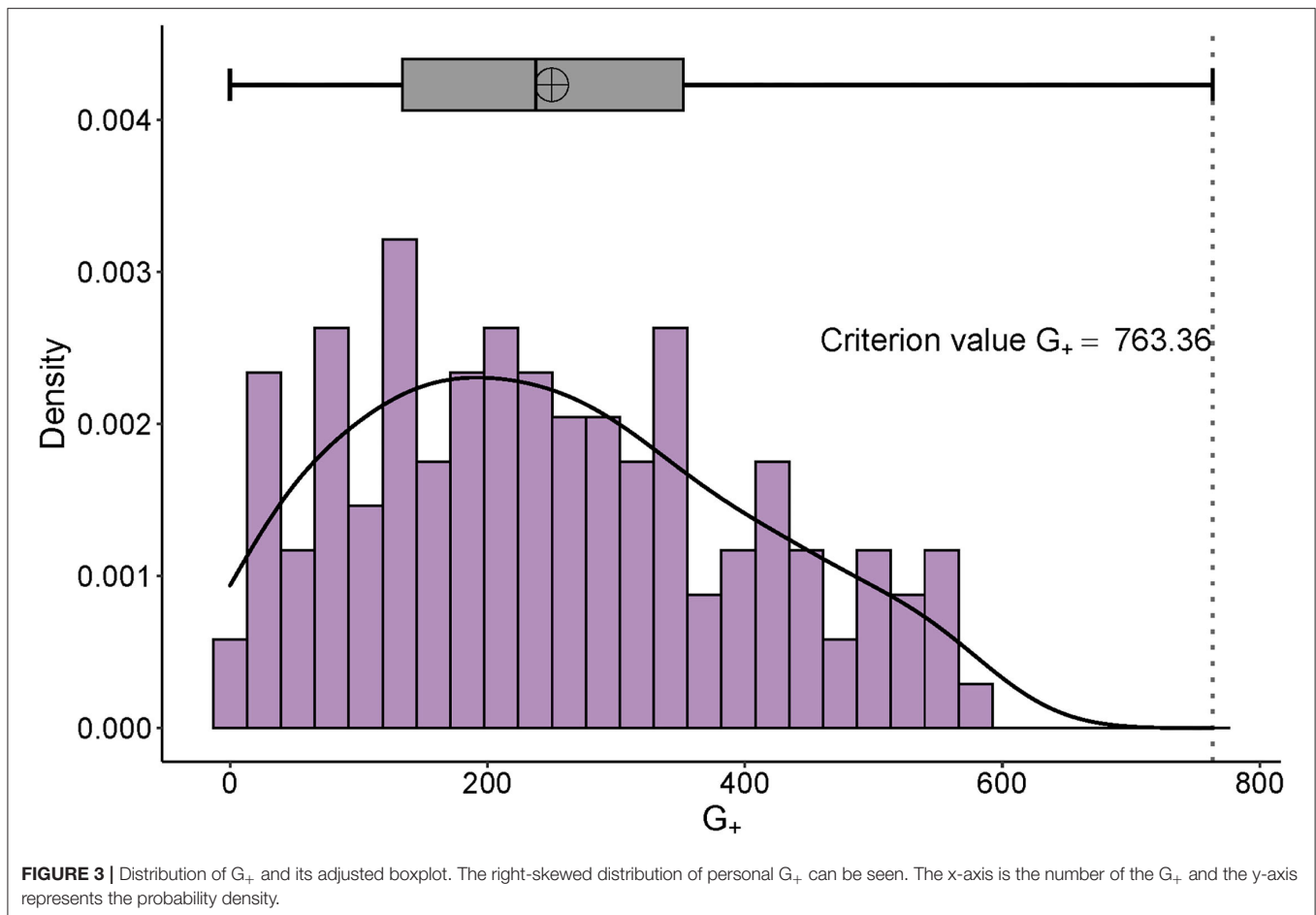
### MSA II: Scale Formation

First, the global scalability coefficient (denoted as *H*) of the items was calculated. It denotes the discrimination power of the items (25). According to Sijtsma and van der Ark (23), if the  $H < 0.3$ , the set is unscalable; if  $0.3 \leq H < 0.4$ , the set is a weak scale; if  $0.4 \leq H < 0.5$ , it is a medium scale; if  $H \geq 0.5$ , the scale is strong.

Second, the automatic item selection procedure (AISP) was applied based on the genetic algorithms (26). This genetic parameter includes the size of sampling items = 20, the cross-over probability = 0.05, and mutation probability = 0.1. The scalability coefficient boundary value starts from 0.3 to 0.54 (step length = 0.03). To meet the minimum sample size requirements for MSA (27), the threshold value was selected 0.42.

**Hypothesis I: Local Independence.** The heat map was based on a pairwise scalability coefficient between items *i* and *j* (denoted as  $H_{ij}$ ). The monotone homogeneity model of the Mokken scale implies that  $0 \leq H_{ij} \leq 1$ . The values outside this range are violations. *W* values were calculated based on conditional association to estimate the degree of a certain item that is suspected regarding the local dependence (28). The extreme values of *W* (denoted as  $W_+$ ) are identified by the Tukey fence algorithm:  $W_+ > Q3 + 3^*(Q3 - M)$ . *M* is the median and *Q3* is the 3rd percentile. The minimum sample size of a rest-score group was 4. No weight was set for the sample size on each conditional covariance. The weight of each conditional covariance on the computation of *W*<sub>1</sub>, *W*<sub>2</sub>, and *W*<sub>3</sub> was the proportion of negative covariances. The minimum sample size of the conditioning variable to compute a covariance was 4.





**Hypothesis II: Monotonicity.** The parameter has set the minimum sample size as 50 and the lowest item response function value as 0.03 (15). Four indices will be reported. (1) #ac: the number of possible violations; (2) #vi: the actual number of violations; (3) #zsig: number of statistically significant violations, 0 indicating no violations; and (4) Crit: the critical value summarizes effect size of violation (29). Although Crişan et al. (29) reported that the Crit has poor power given a small sample size of 100, we decided to report this routine index for more comprehensive considerations.

**Hypothesis III: IIO.** i.e., The items have a fixed rank of difficulty irrespective of the level of personal ability. The function was set as *manifest invariant item ordering (MIIO)* by conducting a backward item selection procedure to make the final decision.

The items violating one of the three hypotheses were removed. The remained items entered the reliability testing step.

### MSA III: Reliability Testing

Four reliability statistics were calculated: Cronbach's  $\alpha$ , Guttman's  $\lambda_2$ , Molenaar Sijtsma  $\rho$ , and the latent class reliability coefficient (LCRC) (25). They all range from 0 to 1. A larger value indicates stronger internal consistency. The Cronbach's  $\alpha$  and the Guttman's  $\lambda_2$  are two traditional coefficients of reliability.

However, they may generate biased estimations for non-parametric models. The Molenaar Sijtsma  $\rho$  is more suitable for scales with IIO. The LCRC provides unbiased estimations for the Mokken scale without limiting the condition of IIO.

### Rasch Modeling Stages

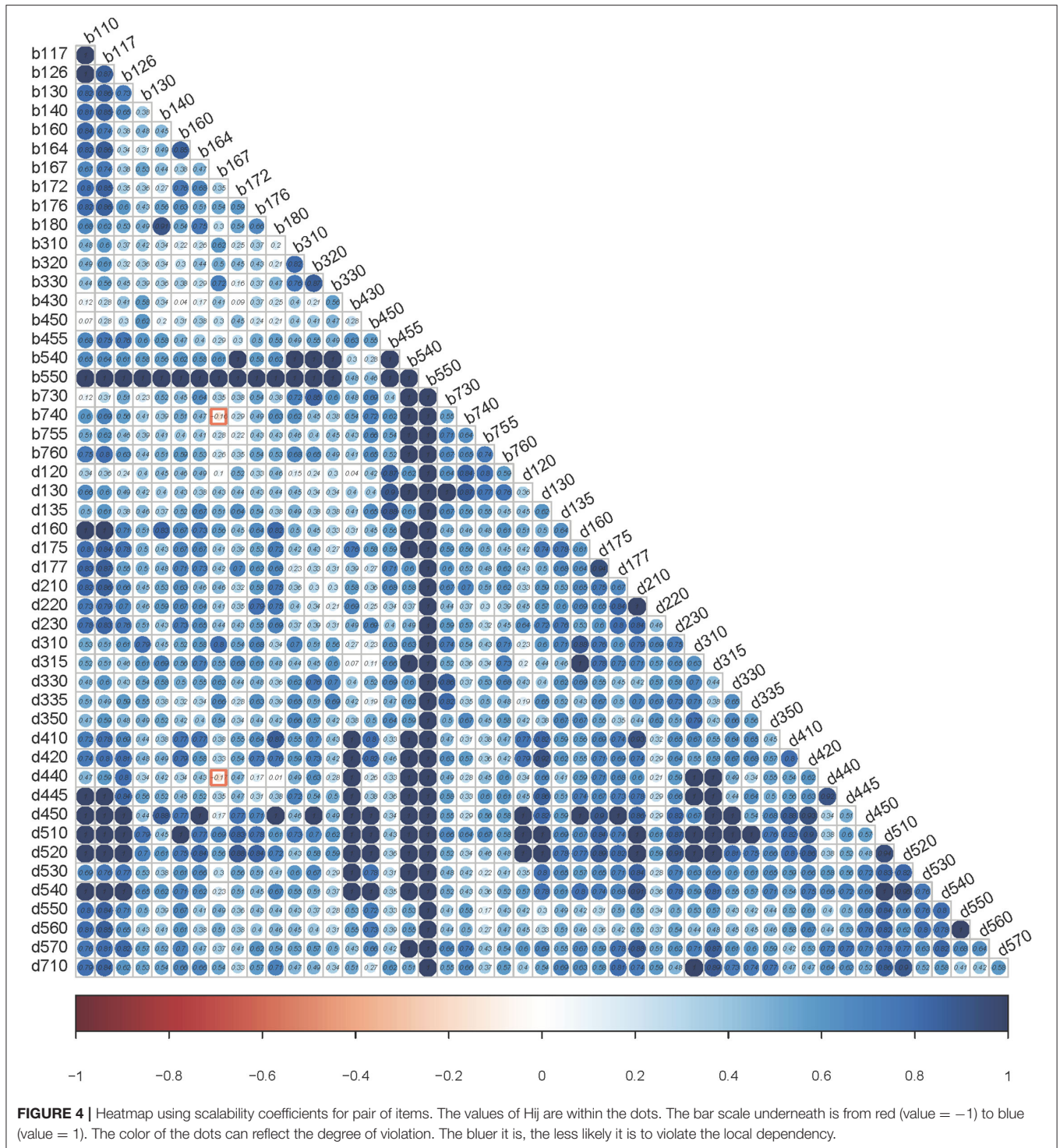
#### Rasch Stage I: Item Screening Circle

First, point biserial correlation with a negative coefficient between each item and the total score (either the total score with or without the certain item score) was all examined. The flagged items were recorded in the list of X1.

Second, an optimal model is selected from the constrained (discrimination = 1) and unconstrained (discrimination  $\neq$  1) Rasch models based on the likelihood ratio test. The model with the lower value of the Akaike information criterion (AIC) is preferred.

Third, the item goodness-of-fit test was performed based on  $\chi^2$  statistics. The items with  $p < 0.05$  were recorded in the list of X2.

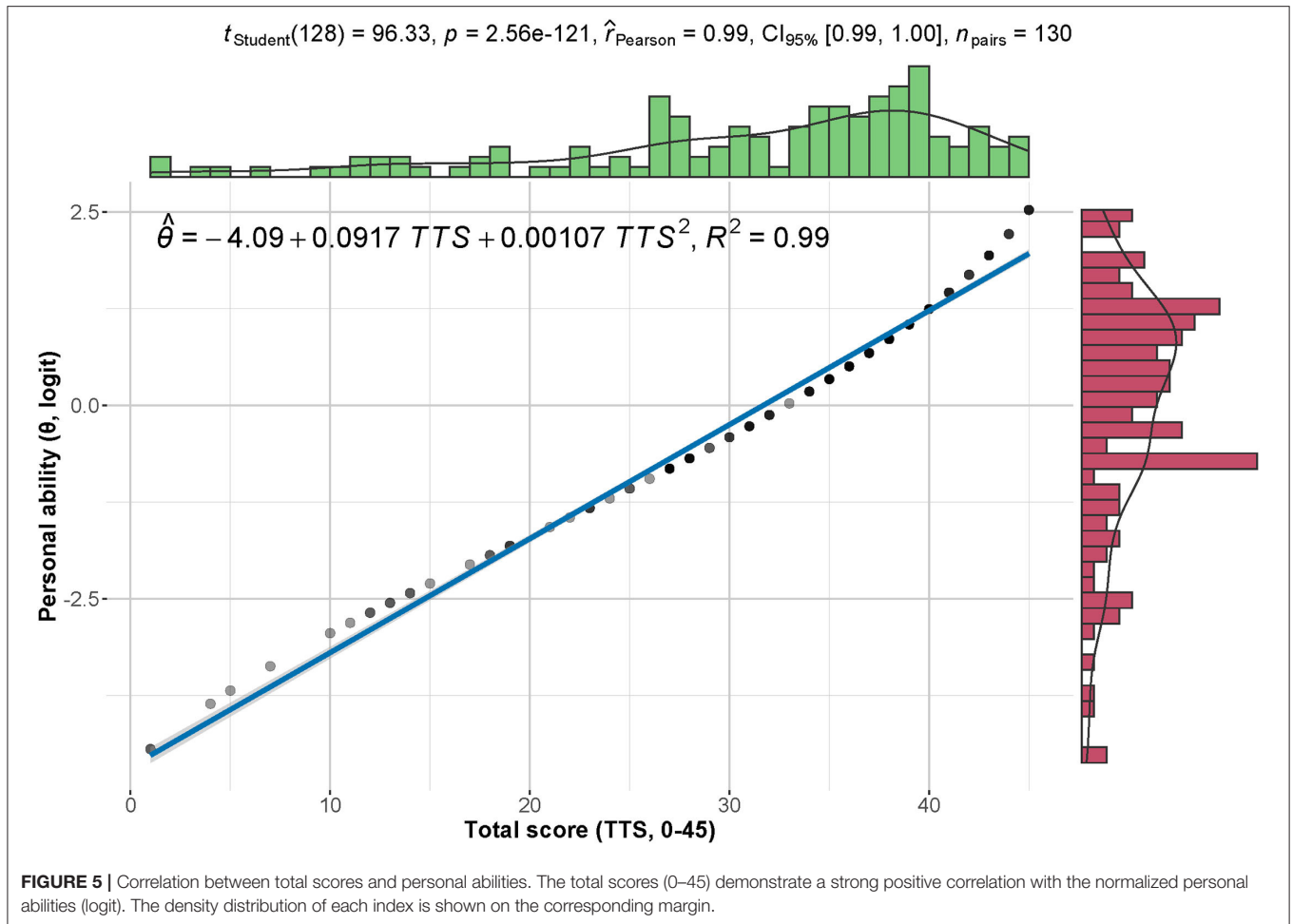
Fourth, if the combination of X1 and X2 is non-empty, the items in it will be removed from the candidate set and the next circle will start. If the combination set is empty, the circle will stop.



**Rasch Stage II: Model Checking**

**Global Goodness-of-Fit Testing.** The global goodness-of-fit of the model was tested by the parametric bootstrap test using Pearson’s  $\chi^2$  statistic. The null hypothesis states that the observed data have been generated under the Rasch model with parameter values of the maximum likelihood estimates  $\hat{\theta}$ . The specific test method

is to simulate the standard Rasch model based on the estimated parameter  $\hat{\theta}$  from the candidate Rasch model. As a result, it produces simulated data sets (number =  $B$ ), each of which can calculate the Pearson  $\chi^2$  value (denoted as  $T_b$ ). The observed data can also calculate the Pearson  $\chi^2$  value (denoted as  $T_{obs}$ ). By calculating the number of  $T_b > T_{obs}$  (denoted as  $N_+$ ), the  $p$ -value



can be achieved by the following equation:  $(1+N_+)/(B+1)$ . If  $p > 0.05$ , it supports  $H_0$ , i.e., the measured values are determined with confidence from the simulated standard Rasch model based on the same parameters of the candidate Rasch model. We set the  $B = 199$  for this procedure.

**Unidimensional Testing.** A total of 100 unidimensional models were built with the Monte Carlo simulation method. The alternative hypothesis was that the second eigenvalue of the observed data is substantially larger than the second eigenvalue of data under the assumed model. If the test shows  $p > 0.05$ , it indicates the candidate model is not significantly different from the simulated unidimensional models.

If the candidate model passes both tests, it becomes the final Rasch model for further parameter estimating.

### Rasch Stage III: Parameters Estimating

**Personal Ability vs. Total Score of Items.** The correlation was assessed between the total score of the items and the normalized personal ability level estimated by the final Rasch model.

**Item Positions.** The item difficulties, i.e., the item positions, were estimated for the final Rasch model. The  $\chi^2$  test was exploited to check the goodness of fit for each item in the model.

The Bonferroni method was used to calibrate the  $p$ -value. Items with adjusted  $p > 0.05$  fit good to the model.

**Item Characteristic Curve (ICC).** Visualization for checking the shape and relation of curves, especially monotonicity and IIO.

**Differential Item Functioning Analysis.** In this study, value 1 stood for “functioning or no disabled” concerning the ICF items. Lord’s  $\chi^2$  analysis embedded in the “difR” package of R was employed to analyze the gender as the only DIF of the model (30). The  $p$ -values of multiple comparisons were adjusted by Holm adjustment.

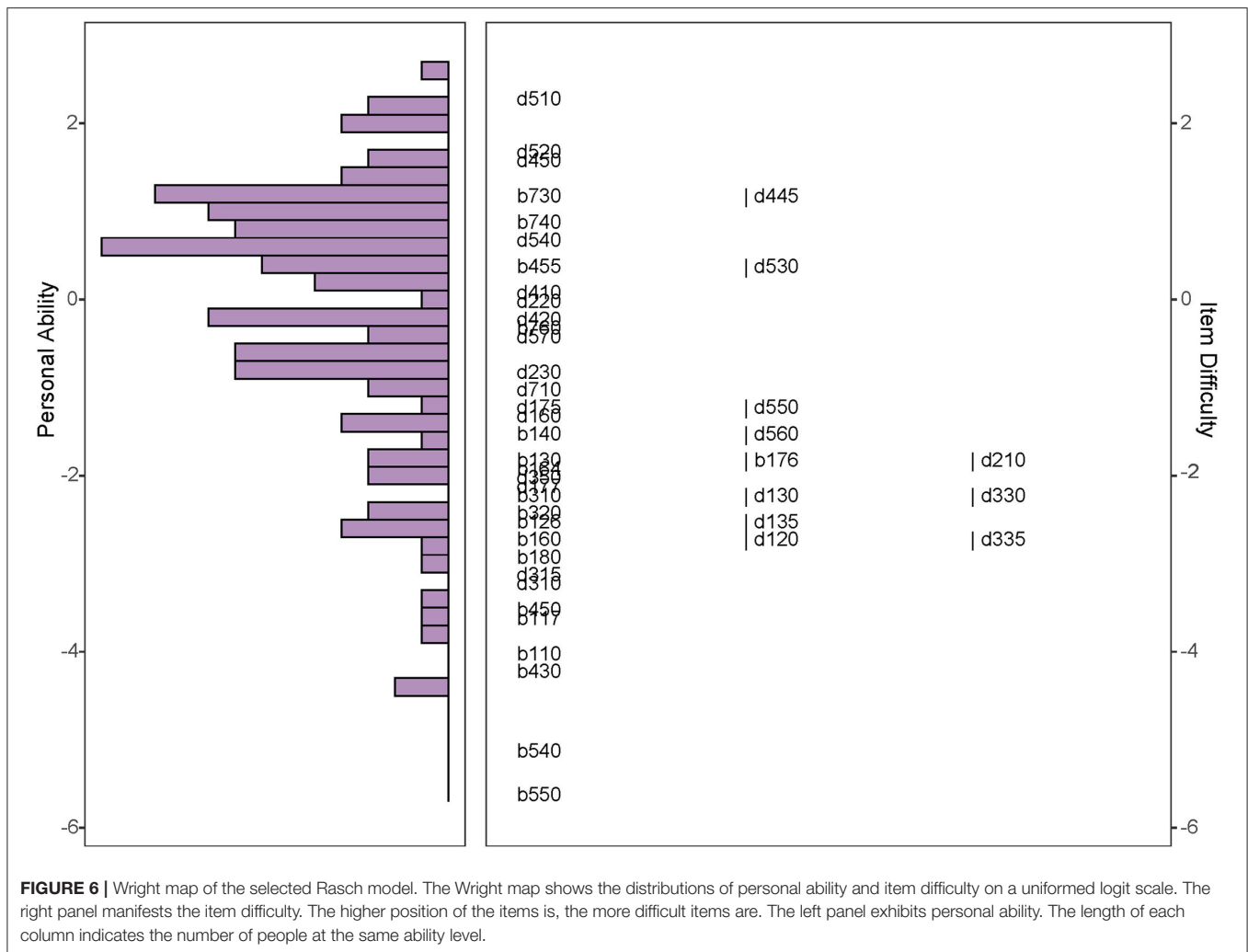
**The Estimated Personal Ability vs. the MBI.** Pearson correlation coefficient was estimated for justification of using the Rasch model to measure functioning levels.

## RESULTS

### Demographics

A total of 130 stroke patients with plegia (36 women and 94 men) were recruited in this research. As described in mean  $\pm$  SD [minimum, maximum], the age was  $64.9 \pm 13$  [28, 87] years. The duration of stroke was  $4.46 \pm 8.62$  [0, 58] months, the MBI was  $47.7 \pm 25.7$  [0, 100] scores, and the body mass index was  $24.3 \pm 3.85$  [15.62, 39.86]  $\text{kg}/\text{m}^2$ . There was no significant difference in





**FIGURE 6 |** Wright map of the selected Rasch model. The Wright map shows the distributions of personal ability and item difficulty on a uniformed logit scale. The right panel manifests the item difficulty. The higher position of the items is, the more difficult items are. The left panel exhibits personal ability. The length of each column indicates the number of people at the same ability level.

education time ( $p = 0.324$ ), solitary status ( $p = 0.774$ ), diagnosis ( $p = 0.475$ ), recurrence ( $p = 1.000$ ), plegia side ( $p = 0.286$ ), age ( $p = 0.1213$ ), the course of disease ( $p = 0.1629$ ), body mass index ( $p = 0.6305$ ), and scores of MRS ( $p = 0.675$ ) and MBI ( $p = 0.6305$ ) across gender. The detailed description of the demographic data was manifested in **Table 1**.

**MSA Results**

**MSA I: Data Shaping**

All participants were below the extreme value and qualified for further analysis (**Figure 3**, adjusted criterion value  $G_+ = 763.36$ ). According to the screening criterion of  $\geq 5\%$  censored data, there were 25 items removed. **Appendix 1** reveals 57 items of b, and 36 items of d category were further analyzed by MSA.

**MSA II: Scale Formation**

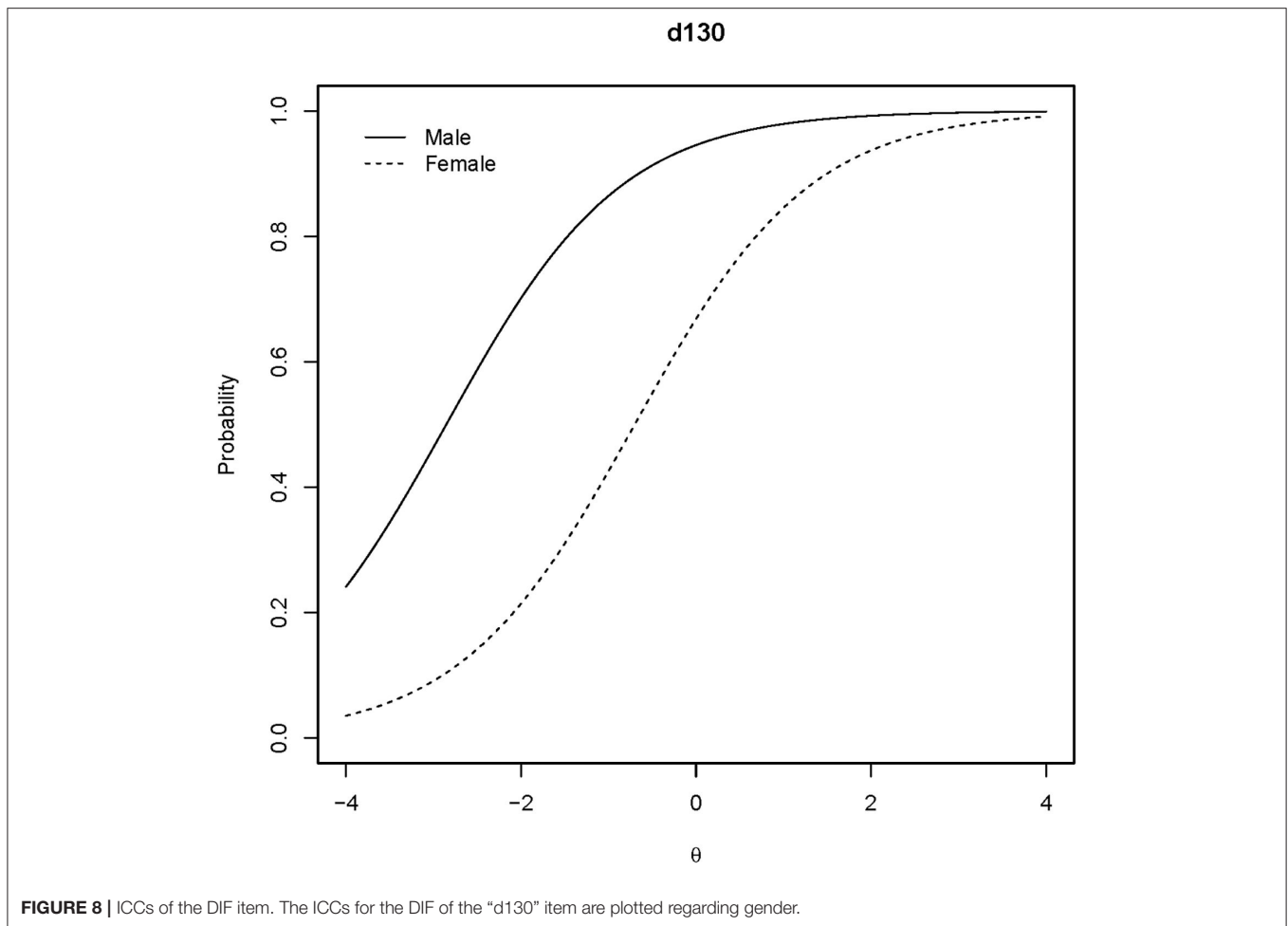
**Global Scalability Coefficient Estimating.** The H of the 93 items = 0.3619 (standard error = 0.0401), which was less accuracy ( $0.3 < H < 0.4$ ) for generating the Mokken scale and ought to be improved by further item selection procedures (23).

**AISP. Appendix 2** listed the outcome patterns of AISP with different cut values of scalability coefficients. Sample size in the range of 50–250 requires at least a cut value of  $c = 0.42(27)$  to retain the predominance of scale 1 and construct a unidimensional model (**Appendix 3**).

**Homogeneity Coefficients Measuring. Appendix 4** demonstrated that the overall homogeneity coefficient was 0.5446 with a standard error of 0.0437. Two things should be noted: according to the standard errors of “b110 Consciousness functions,” “b117 Intellectual functions,” “b180 Experience of self and time functions,” “b430 Hematological system functions,” “b450 Additional respiratory functions,” and “b540 General metabolic functions” more than 0.1, these 6 items demonstrate the relatively low accuracy of the measurement. However, the  $H_i$  of 50 items is all  $> 0.42$ , which suggests that the sample size is suitable for the MSA of this item set.

**Hypothesis I: Local Independence.** The scalability coefficient  $H_{ij}$  is the normed covariance between items  $i$  and  $j$ . Negative values violate the hypotheses of the monotone homogeneity model of the Mokken scale, and positive values tend to but do not necessarily support the monotonicity, local independence,





of fit, we applied a quadratic model for the linear relationship between the total score and the percentage value of normalized personal ability.

**Figure 6** is the Wright map of the final Rasch model that showed the distributions of personal ability and item difficulty among those ICF categories. The negative peak of personal ability revealed that more participants had a relatively high ability. **Appendix 7** and the right panel of **Figure 6** suggest that four items were relatively easy, namely, “b110 Consciousness functions,” “b430 Hematological system functions,” “b540 General metabolic functions,” and “b550 Thermoregulatory functions.”

**Figure 7** is the ICC of the final Rasch model. The “S” shapes and parallel distributions of the curves supported monotonicity and IIO.

**Appendix 8** shows Lord’s  $\chi^2$  test for differential item functioning analysis of gender. The Holm adjusted  $p$ -values suggested that the items are not DIF items of gender except for the “d130 Copying” (Holm adjusted  $p = 0.0261 < 0.05$ ). **Figure 8** shows that the “d130” is a uniformed DIF. The Welch two-sample  $t$ -test for personal abilities considering genders showed  $t_{(64,59)} = -0.3432$  with  $p =$

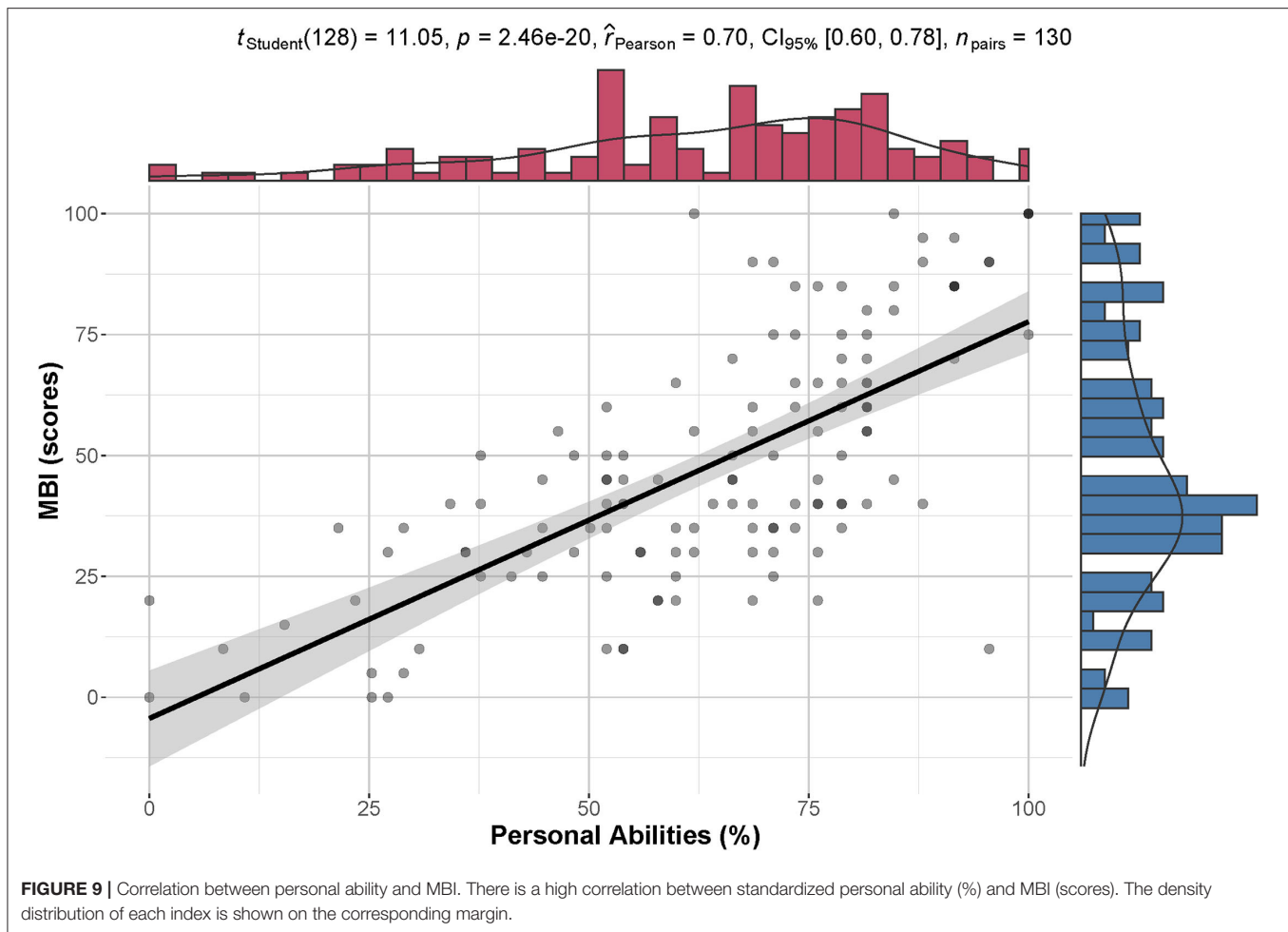
0.7325. The absolute difference in means of personal abilities is 0.1021 logit.

**Figure 9** shows the Pearson correlation between individual ability and MBI score. It suggested that the personal abilities calculated in the model had a strong correlation with MBI ( $p = 2.46 \times 10^{-20} < 0.001$  and effect size = 0.70).

**Table 2** is the final assessment scale containing 45 items. The conversional table between the total scores of 45-ICF items and functional ability in percentile score (0–100%) can be found in **Table 3**.

## DISCUSSION

This study provides research examples (**Supplementary Materials 1, 2**) and application ideas for health assessment based on the IRT. By following the pipeline of MSA and Rasch modeling, it was realized by the simple ICF questionnaire (0 = disabled and 1 = functioning) to complete competence evaluation. The total score of the Mokken scale and the estimated value of the personal functional ability of the Rasch model represented the measure of functioning degree.



Rasch modeling has long been used as an ICF study approach. It was an ideal theoretical model for estimating ordered questionnaires (9, 11, 12, 14) for transforming the ordinal scale into the linear interval scale. In addition, both personal ability and item difficulty can be calculated in a Rasch model (14). MSA was performed as the preliminary screening of the scale content. It can not only reduce the complicated data processing by Rasch but also increase the stability and universality of the scale (23). Rusch et al. (31) utilized MSA to extract scalable items that are compatible with the hypotheses for further parametric modeling. We followed this strategy by applying the MSA filtering before the Rasch modeling that constructs the potential functional assessment scale.

The content of our final scale is similar to the subscale of motor, communication, and cognition in Van de Winckel et al. (14). However, there are differences in the item selection between the Lucerne ICF-based multidisciplinary observation scale (LIMOS) and our scale. The LIMOS only contains activities and participation-related ICF items, while the scale we develop not only involves activity and participation (d) but also body function (b) based on the extended ICF core set for stroke. Like the LIMOS, a higher functional level is with a higher score and

represents less disabled. However, the final scale in our study can also be easily answered by “Yes” or “No” without considering the 5-point Likert scale with numeric rating ranks “no impairment” = 0; “mild impairment” = 1, “moderate impairment” = 2, “severe impairment” = 3, and “complete impairment” = 4.

Our scale has fewer items (45 categories) than the initial b and d items (118 categories) in the extended comprehensive stroke ICF core set but contains extensive items compared to the brief stroke core set (18 categories) (32). Since the yes-or-no response is the simplest task of the questionnaire, our dichotomous model makes it possible to easily embed the scale in clinical practices. It can relieve health practitioners from the burden of multiple assessment scales.

The significantly strong correlations between personal ability and MBI indicated that the scale can measure the activities of daily living. If we dig into the details of the final scale, most of the 10 aspects of MBI were covered by the categories, especially the items with bold fonts in **Appendix 7**. The final 45-item scale included more aspects of daily functioning, for example, “d570 Looking after one’s health,” “d710 Basic interpersonal interactions,” and “b130 Energy and drive functions.” Moreover, the model not only provided an assessment tool but also



**TABLE 2** | The final assessment scale with the Rasch model.

Categories	Category title	Disabled (0)	Functioning (1)
b110	Consciousness functions		
b117	Intellectual functions		
b126	Temperament and personality functions		
b130	Energy and drive functions		
b140	Attention functions		
b160	Thought functions		
b164	Higher-level cognitive functions		
b176	Mental function of sequencing complex movements		
b180	Experience of self and time functions		
b310	Voice functions		
b320	Articulation functions		
b430	Hematological system functions		
b450	Additional respiratory functions		
b455	Exercise tolerance functions		
b540	General metabolic functions		
b550	Thermoregulatory functions		
b730	Muscle power functions		
b740	Muscle endurance functions		
b760	Control of voluntary movement functions		
d120	Other purposeful sensing		
d130	Copying		
d135	Rehearsing		
d160	Focusing attention		
d175	Solving problems		
d177	Making decisions		
d210	Undertaking a single task		
d220	Undertaking multiple tasks		
d230	Carrying out daily routine		
d310	Communicating with—receiving—spoken messages		
d315	Communicating with—receiving—non-verbal messages		
d330	Speaking		
d335	Producing non-verbal messages		
d350	Conversation		
d410	Changing basic body position		
d420	Transferring oneself		
d445	Hand and arm use		
d450	Walking		
d510	Washing oneself		
d520	Caring for body parts		
d530	Toileting		
d540	Dressing		
d550	Eating		
d560	Drinking		
d570	Looking after one's health		
d710	Basic interpersonal interactions		

*The condition of the disabled refers to moderate, severe, or complete dysfunction. The condition healthy refers to no or mild dysfunction.*

offered meaningful insights for intervention. For instance, “d510 Washing oneself,” “d520 Caring for body parts,” and “d450 Walking” are the items with top difficulties. These are essential activities for health experiences and are usually scheduled as main rehabilitation aims.

We should emphasize that the application of the final scale assigns 1 for “functioning” and 0 for “disability.” Rasch model offers two critical values, namely, personal ability and item difficulty. If the disability score = 1, the personal ability means “the disability level of the person,” and the item difficulty implies

**TABLE 3** | The converted form between 45-ICF item scores and functional ability (0–100%).

Total score	$\theta$ (Logit)	Percent	Total score	$\theta$ (Logit)	Percent
0	-4.0889	0	23	-1.4149	43
1	-3.9961	1	24	-1.2731	45
2	-3.9011	3	25	-1.1291	47
3	-3.8041	5	26	-0.983	49
4	-3.7049	6	27	-0.8348	52
5	-3.6035	8	28	-0.6844	54
6	-3.5001	9	29	-0.5319	57
7	-3.3945	11	30	-0.3773	59
8	-3.2867	13	31	-0.2205	62
9	-3.1769	15	32	-0.0616	64
10	-3.0649	16	33	0.0994	67
11	-2.9507	18	34	0.2625	69
12	-2.8345	20	35	0.4278	72
13	-2.7161	22	36	0.5952	75
14	-2.5956	24	37	0.7648	77
15	-2.4729	26	38	0.9365	80
16	-2.3481	28	39	1.1103	83
17	-2.2212	30	40	1.2862	85
18	-2.0922	32	41	1.4643	88
19	-1.961	34	42	1.6445	91
20	-1.8277	36	43	1.8269	94
21	-1.6922	38	44	2.0113	97
22	-1.5546	40	45	2.198	100

Total score (TTS) represents the scores of 45-ICF items. The theta (Logit) can be calculated by the equation  $\hat{\theta} = -4.09 + 0.0917TTS \times 0.00107TTS^2$ . Percent is the normalized personal ability in percentile score (0–100%).

“the difficulty of the item to be dysfunction.” If the functioning score = 1, the personal ability signifies “the functional level of the person,” and the item difficulty indicates “the difficulty of the item to be healthy.” Previous reports emphasize the purpose of estimating the functional level for a person rather than focusing on clinical intervention for a specific ICF item. Therefore, they score functioning as 0 and disability as 1. They did not differentiate the “difficulty of being dysfunction” and the “difficulty of being healthy.” In contrast, if the functioning is scored as 1 and disability as 0, it can provide the physicians and therapists with item difficulty values that are more in line with their intuitive clinical thinking. It also provides information on whether the certain ICF item should be one of the rehabilitation targets (Is the function deficit difficult to address) or where its rank (How hard the ICF item is) should be put on the intervention schedule of several targets.

The sample size is one of the limitations of this study. Although we controlled this limitation by referring to the study of Straat et al. (27) and followed the principle of maximum variation sampling strategy (20), the ideal solution is still to be large sampling. In addition, our study participants were Chinese people. Considering the cultural differences, our study results might capture the characteristics of stroke among the Chinese population and might not be suitable for other cultural backgrounds. The third limitation is the validity of **Table 3**, although we provided evidence of the correlation between

personal ability and MBI, improvements in the future study could include more powerful and widely accepted tools such as the SF-36, functional independence measure, and Fugl-Meyer scale. The fourth limitation is described in our inclusion criterion. All the post-stroke persons were diagnosed with plegia and capable of providing consent. Therefore, regarding the rank of item difficulties, some higher brain functioning categories are relatively easy, such as “d350 Conversation,” “d177 Making decisions,” “b126 Temperament and personality functions,” and “b117 Intellectual functions.” Following the workflow of this study, the model can be furtherly expanded among the post-stroke population without plegia. The fifth aspect that can be improved is other parametric models. Except for difficulty and discrimination as 1, the final scale can be deeply analyzed by multiple parameters such as pseudoguessing and careless responding. By using multidimensional models, more items with scale number 2, 3, or 4 may be included. In this study, only the gender DIF was explored based on our sample size. The DIF values will be further predicted by education time, solitary status, age (below and above 60 years old), diagnostic category, and the duration of disease after recruiting more participants.

However, the primary aim of this study is to provide a simple ICF-based tool for assessing functioning levels in persons with stroke. The resulted Mokken scale and Rasch model are not perfect, but they provide the first step to improve this strategy.

## CONCLUSION

This study dealt with two questions about the ICF application. First, we evaluated the degree of health (functioning) itself rather than focusing on certain dysfunctional conditions (disability). Second, we completed the quantitative assessment of personal abilities for Chinese stroke persons diagnosed with plegia. This study put forward new ideas of calculating individual functioning through the MSA-based Rasch model. The 45-item scale generated from MSA and Rasch analysis can be an assessment tool for potential functional competence. Moreover, the final scale can guide the grading of individual functioning levels in the process of stroke diagnosis and treatment.

## DATA AVAILABILITY STATEMENT

The datasets presented in this article are readily available. Requests to access the datasets should be directed to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of The First Rehabilitation Hospital of Shanghai (IRB# YK-2021-02-011) and The Affiliated Sir Run Run Hospital of Nanjing Medical University (IRB#

2018-SR-017). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

CF designed the protocol, recruited subjects, and wrote the manuscript. FL contributed to the research concept, supervised the entire study, proofread the manuscript, performed the analysis, and generated the images. ZLJ contributed to the research concept and proofread the manuscript. MXS helped recruit the subjects and gave useful suggestions. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We would love to thank Amanda Ferland, Doctor of Physical Therapy, and Dr. Shou-Guo Liu for helping our proofreading and providing suggestions from clinical experts' opinions. We appreciate Prof. Jia-Yuan Yu for help in reviewing our methods and providing useful and valuable criticisms to the final version of the manuscript. The authors also appreciate all the study's participants' cooperation.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2022.827247/full#supplementary-material>

## REFERENCES

- Vanbellingen T, Ottiger B, Pflugshaupt T, Mehrholz J, Bohlhalter S, Nef T, et al. The responsiveness of the lucerne ICF-based multidisciplinary observation scale: a comparison with the functional independence measure and the barthel index. *Front Neurol.* (2016) 7:152. doi: 10.3389/fneur.2016.00152
- Lins L, Carvalho FM. SF-36 total score as a single measure of health-related quality of life: scoping review. *SAGE Open Med.* (2016) 4::2050312116671725. doi: 10.1177/2050312116671725
- Liu M, Chino N, Tuji T, Masakado Y, Hase K, Kimura A. Psychometric properties of the stroke impairment assessment set (SIAS). *Neurorehabil Neural Repair.* (2002) 16:339–51. doi: 10.1177/0888439002239279
- Castro HHG, Alencar AP, Benseñor IM, Lotufo PA, Goulart AC. Multimorbidities are associated to lower survival in ischaemic stroke: results from a brazilian stroke cohort (EMMA Study). *Cerebrovasc Dis Basel Switz.* (2017) 44:232–9. doi: 10.1159/000479827
- Kapoor A, Lanctôt KL, Bayley M, Kiss A, Herrmann N, Murray BJ, et al. “Good outcome” isn’t good enough. *Stroke.* (2017) 48:1688–90. doi: 10.1161/STROKEAHA.117.016728
- World Health Organization. *How to Use the ICF: A Practical Manual for Using the International Classification of Functioning, Disability and Health (ICF). Exposure Draft for Comment [Internet]* (2013). Available online at: <https://www.medbox.org/document/how-to-use-the-icf-a-practical-manual-for-using-the-international-classification-of-functioning-disability-and-health-icf-exposure-draft-for-comment#GO> (accessed July 30, 2021).
- Rauch A, Cieza A, Stucki G. How to apply the international classification of functioning, disability and health (ICF) for rehabilitation management in clinical practice. *Eur J Phys Rehabil Med.* (2008) 44:329–42.
- Jette AM. The Utility of and need for improving the ICF. *Phys Ther.* (2018) 98:629–30. doi: 10.1093/ptj/pzy056
- Ehrmann C, Proding B, Stucki G, Cai W, Zhang X, Liu S, et al. ICF generic set as new standard for the system wide assessment of functioning in China: a multicentre prospective study on metric properties and responsiveness applying item response theory. *BMJ Open.* (2018) 8:e021696. doi: 10.1136/bmjopen-2018-021696
- Liu S, Reinhardt JD, Zhang X, Ehrmann C, Cai W, Proding B, et al. System-wide clinical assessment of functioning based on the international classification of functioning, disability and health in China: interrater reliability, convergent, known group, and predictive validity of the ICF Generic-6. *Arch Phys Med Rehabil.* (2019) 100:1450–7.e1. doi: 10.1016/j.apmr.2018.11.014
- Gao Y, Yan T, You L, Li K, Zhang L, Zhang M. Psychometric properties of the international classification of functioning, disability and health rehabilitation set: a rasch analysis. *Int J Rehabil Res.* (2020) 44:144–51. doi: 10.1097/MRR.0000000000000463
- Li K, Yan T, You L, Xie S, Li Y, Tang J, et al. psychometric properties of the international classification of functioning, disability and health set for spinal cord injury nursing based on rasch analysis. *Disabil Rehabil.* (2018) 40:338–45. doi: 10.1080/09638288.2016.1250169
- Jia M, Tang J, Xie S, He X, Wang Y, Liu T, et al. Using a mobile app-based international classification of functioning, disability, and health set to assess the functioning of spinal cord injury patients: rasch analysis. *JMIR MHealth UHealth.* (2020) 8:e20723. doi: 10.2196/20723
- Van de Winckel A, Ottiger B, Bohlhalter S, Nyffeler T, Vanbellingen T. Comprehensive ADL outcome measurement after stroke: rasch validation of the lucerne ICF-based multidisciplinary observation scale (LIMOS). *Arch Phys Med Rehabil.* (2019) 100:2314–23. doi: 10.1016/j.apmr.2019.02.012
- Stochl J, Jones PB, Croudace TJ. Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric irt method

- in empirical research for applied health researchers. *BMC Med Res Methodol.* (2012) 12:74. doi: 10.1186/1471-2288-12-74
16. Vaughan B, Grace S. A Mokken scale analysis of the peer physical examination questionnaire. *Chiropr Man Ther.* (2018) 26:6. doi: 10.1186/s12998-018-0176-0
  17. Zhang L, Li Z. A Mokken scale analysis of the kessler-6 screening measure among chinese older population: findings from a national survey. *BMC Geriatr.* (2020) 20:361. doi: 10.1186/s12877-020-01771-w
  18. Kuijpers RE, van der Ark LA, Croon MA, Sijtsma K. Bias in point estimates and standard errors of mokken's scalability coefficients. *Appl Psychol Meas.* (2016) 40:331–45. doi: 10.1177/0146621616638500
  19. Koopman L, Zijlstra BJH, van der Ark LA. A two-step, test-guided mokken scale analysis, for nonclustered and clustered data. *Qual Life Res.* (2021) 31:25–36. doi: 10.1007/s11136-021-02840-2
  20. Patton MQ. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice.* Saint Paul, MN: SAGE Publications (2014).
  21. Starrost K, Geyh S, Trautwein A, Grunow J, Ceballos-Baumann A, Prosiegel M, et al. Interrater reliability of the extended ICF core set for stroke applied by physical therapists. *Phys Ther.* (2008) 88:841–51. doi: 10.2522/ptj.20070211
  22. Newgard CD, Haukoos JS. Advanced statistics: missing data in clinical research—part 2: multiple imputation. *Acad Emerg Med Off J Soc Acad Emerg Med.* (2007) 14:669–78. doi: 10.1111/j.1553-2712.2007.tb01856.x
  23. Sijtsma K, van der Ark LA. A tutorial on how to do a mokken scale analysis on your test and questionnaire data. *Br J Math Stat Psychol.* (2017) 70:137–58. doi: 10.1111/bmsp.12078
  24. Rizopoulos D. Ltm: an R package for latent variable modeling and item response analysis. *J Stat Softw.* (2006) 17:1–25. doi: 10.18637/jss.v017.i05
  25. Sengul Avsar A, Tavsanlı E. Examination of polytomous items' psychometric properties according to nonparametric item response theory models in different test conditions. *Educ Sci Theory Pract.* (2017) 17:493–514. doi: 10.12738/estp.2017.2.0246
  26. Straat JH, van der Ark LA, Sijtsma K. Comparing optimization algorithms for item selection in mokken scale analysis. *J Classif.* (2013) 30:75–99. doi: 10.1007/s00357-013-9122-y
  27. Straat JH, van der Ark LA, Sijtsma K. Minimum sample size requirements for mokken scale analysis. *Educ Psychol Meas.* (2014) 74:809–22. doi: 10.1177/0013164414529793
  28. Straat JH, van der Ark LA, Sijtsma K. Using conditional association to identify locally independent item sets. *Methodology.* (2016) 12: 117–23.
  29. Crişan D-R, Tendeiro J, Meijer R. The Crit Value as an Effect Size Measure for Violations of Model Assumptions in Mokken Scale Analysis for Binary Data [Internet] (2019). Available from: <https://psyarxiv.com/8ydmr/> (accessed August 6, 2021).
  30. Magis D, Béland S, Tuerlinckx F, De Boeck P. A general framework and an R package for the detection of dichotomous differential item functioning. *Behav Res Methods.* (2010) 42:847–62. doi: 10.3758/BRM.42.3.847
  31. Rusch T, Mair P, Hatzinger R. *Psychometrics With R: A Review of CRAN Packages for Item Response Theory.* (2013). Available online at: <https://core.ac.uk/display/18450228> (accessed August 4, 2021).
  32. Geyh S, Cieza A, Schouten J, Dickson H, Frommelt P, Omar Z, et al. ICF core sets for stroke. *J Rehabil Med.* (2004) (44 Suppl):135–41. doi: 10.1080/16501960410016776
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Feng, Jiang, Sun and Lin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.