European Journal of Physical and Rehabilitation Medicine
EDIZIONI MINERVA MEDICA

Evaluation of manual ability in stroke patients in Benin: cultural adaptation and Rasch validation of the ABILHAND-Stroke questionnaire

Ditouah Didier NIAMA NATTA, Charles Sebiyo BATCHO, Gaetan STOQUART, Etienne ALAGNIDé, Toussaint KPADONOU, Thierry LEJEUNE

European Journal of Physical and Rehabilitation Medicine 2018 Aug 29
DOI: 10.23736/S1973-9087.18.05195-X

Article type: Original Article

Article first published online: August 29, 2018
Manuscript accepted: August 27, 2018
Manuscript revised: July 25, 2018
Manuscript received: January 22, 2018

Evaluation of manual ability in stroke patients in Benin: cultural adaptation and Rasch validation of the ABILHAND-Stroke questionnaire

Ditouah Didier Niama Natta [1,2], Charles Sebiyo Batcho [3,4], Gaëtan G. Stoquart [2,5,6], Etienne Alagnidé [1], Toussaint Kpadonou [1], Thierry M. Lejeune [2,5,6] *

[1] National University Hospital of Cotonou, Physical Medicine and Rehabilitation Department, Cotonou, Benin, [2] Université catholique de Louvain, Institute of Experimental and Clinical Research, Neuro Musculo Skeletal Lab, Brussels, Belgium, [3] Centre Interdisciplinaire de Recherche en Réadaptation et Intégration Sociale (CIRRIS), Québec, Canada, [4] Department of Rehabilitation, Faculty of Medicine, Université Laval, Québec, Canada, [5] Cliniques universitaires Saint-Luc, Physical Medicine and Rehabilitation Department, Brussels, Belgium, [6] Université catholique de Louvain, Louvain Bionics, Louvain-La-Neuve, Belgium,

*Corresponding author: Thierry Lejeune

Pr Thierry Lejeune, MD PhD

Médecine Physique et Réadaptation

Cliniques universitaires Saint-Luc

Av. Hippocrate 10

1200 Bruxelles, Belgium. E-mail : thierry.lejeune@uclouvain.be

ABSTRACT

Background: ABILHAND is a self-reported questionnaire assessing manual ability. It was validated and calibrated using the Rasch analysis for European stroke patients. After a stroke, performing upper limb activities of daily living is influenced by personal and environmental contextual factors. It is thus important to conduct a contextual validation to use this questionnaire outside of Europe.

Aim: To perform a cross cultural validation of the ABILHAND-Stroke questionnaire for post-stroke patients living in Benin, a West-African country.

Design: Observational cross-sectional study

Setting: Outpatient rehabilitation centres

Population: 223 Beninese chronic stroke patients

Method: The experimental questionnaire was made of 59 items evaluating manual activities. Patients had to estimate their difficulty of performing each activity according to four response categories: impossible, very difficult, difficult and easy. For construct validity analysis, patients were also evaluated with other assessment tools: Box and Block Test, the motor subscale of the Functional Independence Measure, the Stroke Impairment Assessment Set, and ACTIVLIM-Stroke. Data were analysed with the Rasch partial credit model.

Results: The response categories very difficult and difficult were merged and the number of response categories was reduced from 4 to 3 (impossible, difficult and easy). The Rasch analyses selected 16 bimanual activities that fit the Rasch model (chi square=42.35; p=0.10). The item location ranged from -1.10 to 2.24 logits. The standard error ranged from 0.15 to 0.22 logits. There is no differential item functioning between subgroups (age, sex, dexterity, affected side, time since stroke). The person separation index is 0.82. The questionnaire can measure 3 levels of manual ability, similarly to the occidental version.

Conclusion:

The ABILHAND-stroke is a Rasch validated, unidimensional and invariant questionnaire to assess manual ability among Beninese patients. The ordinal score can be transformed into linear score using a conversion table.

Clinical Rehabilitation Impact:

This assessment tool is clinically relevant in Benin, a developing country, since it requires no specific equipment or training. It should promote and standardize assessments for stroke patients in clinical practice and research in this African country.

Keywords: outcome-assessment, questionnaire, Rasch, arm, activity, stroke

Evaluation of manual ability in stroke patients in Benin: cultural adaptation and Rasch validation of the ABILHAND-Stroke questionnaire

Introduction

Stroke is the leading cause of neurological disabilities in adults (1). Upper limb function is more frequently affected, and is harder to restore than lower limb function (1). Four years post stroke, 45% of patients still present some functional upper limb impairments (2) that bear a negative impact on autonomy and quality of life (3).

Accurate, reliable and valid functional evaluations are essential in order to orientate treatments and assess their efficacy (4). The World Health Organization (WHO) developed the International Classification of Functioning, Disability and Health (ICF) to provide a common framework and standardized terminology (5). The ICF activity domain is essential in neurological rehabilitation, since its main objective is to improve patients' functional abilities. Activities' assessment is mainly based on questionnaires to determine what type of activities of daily living (ADL) patients can perform. Several tools are available to assess upper limb activity, such as the Action Research Arm Test, the Wolf Motor Function Test, the Arm Motor Ability Test or ABILHAND (6). The choice of an appropriate assessment tool should be based on its psychometric properties and clinical relevance (time taken to administer the test and interpret its scores, cost, need for specific equipment and/or training, as well as how easily the test can be used in different settings) (7).

The ABILHAND questionnaire was developed to evaluate manual ability within the ICF activity domain (5). Manual ability is defined as the "ability of a person to perform ADL requiring the use of upper limbs" (8). The ABILHAND questionnaire was validated using the Rasch analysis for stroke patients (8) and patients with other pathologies (9). It specifically evaluates bimanual activities (8). It exhibits excellent psychometric properties and is now recommended for assessment of upper extremity activity in research and clinical praxis (10). This assessment tool would be clinically relevant in developing countries, since it requires no specific equipment or training. However, like most assessment tools, it was mainly validated and calibrated for European populations. Recently, a cross-cultural validation was conducted for Brazilian populations (11).

In light of limited resources in developing countries, patients' rehabilitation and functional assessment must be as cost-effective as possible, for the patient as well as the healthcare system. Questionnaires are thus relevant to evaluate patients in these countries. However, to our knowledge, no questionnaire assessing upper limb activity has been validated for the sub-Saharan population. After a stroke, performing upper limb ADL is influenced by personal and environmental contextual factors according to the ICF model (5). It is thus important to conduct a contextual validation to use this questionnaire outside of Europe. The objective of this study was to validate the ABILHAND questionnaire in a population of stroke patients living in Benin, a West-African Sub-Saharan country.

Materials and Methods

The study was approved by the Ethics Committee of the University Hospital National Centre, Hubert Koutoukou Maga (CNHU-HKM) of Cotonou.

An observational cross-sectional study was performed. Classical and modern measurement theories were used to assess the questionnaire psychometric qualities following the COSMIN taxonomy (12).

Patients

A convenience sample of patients was recruited within the outpatient rehabilitation clinics in the Physical Medicine and Rehabilitation department of the CNHU-HKM, and in the Army Hospital of Cotonou between September 2012 to April 2016. In order to be included in this study patients had to meet the following criteria: 18 years of age or older, time since stroke > 6 months, living at home in Cotonou, and autonomous gait (8), i.e. a score of at least 5/7 for the items “toilet transfer” and “locomotion” on the Functional Independence Measure (FIM) (13). Patients were excluded if they presented with the following criteria: major neglect (Bells test > 26/35) (14), any concomitant pathology that could potentially limit manual ability, comprehension and expression aphasia or cognitive disorders (score on the Mini-Mental State Examination < 20) (15).

Development of ABILHAND-Stroke Benin questionnaire (Content and cross-cultural validity)

Based on the literature review of existing scales and experts’ opinions, Penta et al (8) developed an initial 56-item French questionnaire covering various manual activities. This version was submitted to one Beninese stroke patient and to 7 Beninese rehabilitation professionals (six physical therapists and one Physical Medicine and Rehabilitation specialist). They studied its content and cross-cultural validity according to a consensus-based decision-making. Experts were asked to identify items relevant in the socio-cultural context of Benin, to modify items as needed or to suggest other items. Among the initial 56 items, two items were discarded because they did not match the socio-cultural realities in Benin: "tearing open a pack of chips" and “preparing crepe batter”. Four items were modified: “spreading butter on a slice of bread” became “buttering a slice of bread”; “eating a sandwich” became “eating a sandwich (bread with filling)”; “cracking nuts” became “cracking nuts (or palm nuts)”; “opening a letter” became “opening a letter (closed envelope)”. Furthermore, five items were added: “using a matchstick to light a fire”, “praying with a rosary”, “tying a loincloth around the waist”, “using a toothpick to clean your teeth” and “shake someone’s hand”. Finally, the experimental questionnaire administered to patients was made of 59 items (34 unimanual activities and 25 bimanual activities). These experts also classified bimanual items into three difficulty levels: easy bimanual tasks (group A) were those that could be broken down into unimanual sequences; moderately difficult bimanual tasks (group B) i.e. those requiring the use of the healthy hand and the affected hand to stabilize an object; and difficult bimanual tasks (group C) i.e. those requiring the use of the healthy hand combined with a digital activity of the affected hand. The content validity was also studied by analyzing the Spearman correlation between items’ difficulty levels assessed by the experts and the items’ difficulties resulting from Rasch analyses.

Testing procedure

The first part consisted in collecting socio-demographics and functional data as well as other variables: Stroke Impairment Assessment Set (SIAS), an ordinal scale evaluating stroke-related neurological impairments (score 0 to 76) (16); Box and Block Test (BBT) evaluating gross manual dexterity (17); Geriatric Depression Scale (GDS) an ordinal scale assessing depression in elderly individuals (normal ≤9, moderate depression between 10 and 19, severe depression between 20 and 30) (18); motor subscale of the Functional Independence Measure (FIM-Mo) consisting of the first 13 items (score 13 to 91) (13); and ACTIVLIM-Stroke, a Rasch validated questionnaire assessing limitations in performing ADL (19).

The second part of the evaluation consisted in patients completing the French version of the experimental ABILHAND-Stroke Benin questionnaire. They had to self-assess their ability to perform each of the 59 activities using a four-level scale: impossible (0), very difficult (1), difficult (2) and easy (3). Activities not performed in the past 3 months were qualified as “no answer given”. If a patient was unable to speak or read French properly, the assessor performed an interview in French or in the patient’s native language. Indeed, even though French is the official Beninese language, there are around 70 dialects spoken in Benin.

Rasch analyses

The Rasch model has been widely used for the development and validation of assessment tools in rehabilitation (20). Developed by George Rasch (21), this probabilistic model helps build a continuous, linear and unidimensional scale (22). These psychometric properties contribute to the structural validity, which is one component of construct validity (12). The likelihood test has been performed and the result was significant (chi-square=104,66; p<0.0001) then the partial credit model has been used.

Item selection

Items selection was based on the following criteria that were successively checked.

(a) Missing answers: items with a high proportion of missing answers (> 50%) were discarded, since they were considered as unfamiliar or unrepresentative activities for most patients.

(b) Ordering of response categories: patients had to answer the items according to four response categories. Adjacent response categories were separated by a threshold, corresponding to a location where a patient had an equal probability of scoring one of the adjacent categories. When subjects could not discriminate between the different categories, thresholds were disordered (23). For disordered thresholds, the chosen solution was to combine response categories (24).

(c) Scale targeting: a questionnaire is meant to target the specific population it is supposed to assess. Distributions of individual ability levels and item difficulties were plotted on a common metric logit scale to compare the location of persons and items. It allows to evaluate graphically if items difficulties cover all levels of individual abilities (25; 26). Items identified as “easy” or “impossible” by all subjects were removed as useless extreme items. Floor and ceiling effects were assessed by computing the percentage of patients with a manual ability lower than the easiest item or greater than the most difficult item. This percent must be lower than 15% (27).

(d) Item-trait interaction: overall data fit is verified in term of item-trait interaction assessed by a chi-square test. This chi-square test must be non-significant (26).

(e) Individual item and person fit to the Rasch model: two fit statistics were computed as a function of all individual responses to all items, the residual and chi-square. The residual is the difference between the model's expected scores and the observed scores. The standardized residual was obtained by dividing the residual by an estimate of its standard deviation. The standardized residual consisted in a normal distribution with a mean of 0 and a standard deviation of 1. Persons and items presenting with fit residuals located within a range between -2.5 and 2.5 were deemed acceptable (26; 28). The software divided patients into three incremental class intervals according to their level of ability. The chi-square fit index was calculated as the standardized residual sum of the squares for each class interval. A significant Bonferroni adjusted p value meant the items did not fit the Rasch model (29).

(f) Local independence means there was no major correlation between two items after discarding the underlying trait effect. The correlation matrix analysis was used to verify this criterion. When two items had a correlation index ≥ 0.3, the item less contributive to the global fit index was removed (25).

(g) Unidimensionality is the property of the questionnaire to measure a single attribute of the underlying trait. The unidimensional nature of the questionnaire was tested by the method proposed by Smith (25). After a principal component analysis, the factor loadings of the first component' residuals were used to determine two subsets of items that were then compared by a paired t test (25). The percentage of significant tests ($p < 0.05$) should be below 5 % for a questionnaire to be called unidimensional.

(h) Differential item functioning (DIF). If an item fits the Rasch model, patients' responses should only be determined by the subject's level of manual ability and the difficulty level of the activity measured by the item. (21). If the response to an item is influenced by another variable, it means that persons with the same ability level give a different response to this item. So responses to this item are biased and the item exhibits DIF (30). DIF was evaluated with the Wright and Stone method (31) via five variables. The patients were divided in two subgroups based on their sex, affected side; on their median value of age (55 years), time since stroke (2 years), and BBT score (39 blocks). Item location was determined in each subgroup. A Student's t-test and an Intra Class Correlation (ICC) were used to compare items locations between subgroups and for each variable.

Reliability index PSI

Reliability was evaluated in terms of person separation index (PSI) (32). It was calculated as the ratio of the unbiased estimate of the sample standard deviation to the root mean square measurement error of the sample. It can estimate the number of ability levels that can be measured by the questionnaire. A reliability ≥ 0.80 and number of ability levels ≥2 are deemed acceptable (22; 32).

Convergent Validity

Construct validity was also investigated in terms of convergent validity with classic psychometric methods. The convergent validity was studied by analyzing the correlation between the manual ability assessed with ABILHAND-Stroke Benin and other functional ability scales: ACTIVLIM-Stroke, FIM-Mo, SIAS, BBS and GDS. Pearson correlation coefficient was used for continuous variables and Spearman's rank correlation coefficient for ordinal variables. We expected patients with a high score on the ABILHAND-Stroke Benin to also have a high score on the ACTIVLIM-Stroke, FIM-Mo, BBT and SIAS tests. Conversely, as suggested by Penta et al (8), patients with major depression should have a low manual ability score. The correlation index was considered strong when ≥ 0.7; moderate when comprised between 0.3 and 0.7 and weak when < 0.3 (33). Three-way analysis of variance (3-way ANOVA) was used to evaluate the differences in the manual ability score of patients according to the following characteristics: age, sex and affected side. The difficulties of the items retained in both the Beninese and the European ABILHAND questionnaires were compared using a Student's t-test and an ICC.

Results

We recruited 233 patients who fitted the criteria listed above. Mean age was 54 ± 9.7 years. Sample's characteristics are summed up in table 1.

Selection of items

No item was deleted because of a missing response. The first analysis revealed that all 59 items showed disordered thresholds with the categories "very difficult" and "difficult" being not well differentiated by patients (Figure 1A). Thus, these two categories were merged, leading to a three-level scale: impossible (0), difficult (1) and easy (2). After a new analysis, the response categories were properly ordered (Figure 1B).

Then the targeting of the scale was verified. It appeared that the 34 unimanual activities were labelled as easy by all patients. These items did not permit to differentiate patients according to their manual ability level and were discarded. After several successive analyses on the remaining 25 items, four items were discarded because of their poor fit, and five other items were removed because of local dependency (correlation index ≥ 0.3 with other items).

Fit statistics

The final questionnaire included 16 items (Table 2). The overall fit statistics (mean±SD) for items was -0.12±0.61 logit and for persons was -0.27±1.03 logit where fit statistics are standardized to a mean of 0 and a SD of 1. The item-trait interaction showed a good global fit to the model for the scale (chi-square = 42.35; degrees of freedom = 32; p = 0.10). The item location ranged from -1.10 to 2.24 logits, corresponding to an interval of 3.34 logits with a mean item location of 0.00±1.04 logit. The standard error rate varied from 0.15 to 0.22 logit. The thresholds for the different response categories ranged from -2.23 to 3.03 logits. The person location ranged from -2.13 to 3.78 logits, corresponding to an interval of 5.91 logits with a mean person location of 1.0±1.37 logits.

In our sample, none of the patients reported all items impossible (0% floor effect). Nine subjects in our sample had ability greater than 3.03 logits and could answer "easy" to all items (6.4% ceiling effect). Figure 2 shows that the items difficulties cover relatively well the patients' manual abilities, even if the individual ability was overall superior to the difficulty of the items (Figure 2). Table 2 reports items calibration and fit indices. Table 3 underlines the relationship between the total raw scores, the scores in logits and the scores in percentages. In clinical practice, if patient answered all questions, the ordinal score can be easily transformed in interval score, expressed in percentage. If the patient did not answerer all questions, then a software or a website (for instance http://www.rehab-scales.org) is required to transform the ordinal score into interval score. "

Differential item functioning

Item DIF was evaluated through five variables. The Student-t test showed no significant difference in items' difficulty between each factor' subgroup (p>0.05), and ICCs ranged from 0.64 to 0.85 (p<0.001) (Figure 3).

Unidimensionality

The paired t-test significance level (4.52%) was lower than the acceptable 5%. This validates the unidimensional nature of the scale.

Reliability

The ABILHAND-Stroke Benin questionnaire showed good reliability. The PSI was 0.82 after discarding extreme scores. This indicates that three levels of ability can be distinguished within the patient' sample.

Content validity

The items' difficulty levels assessed by the experts are strongly correlated with the items' difficulties resulting from Rasch analyses (rho = 0.88, p<0.001).

Convergent validity

The patients' manual ability evaluated with ABILHAND-Stroke was moderately to strongly correlated with their neurological impairments, manual dexterity, and ability to perform ADL ($p<0.001$, rho and r ranged from 0.56 to 0.89, Table 4). On the other hand, their manual capacity was negatively, albeit discretely, correlated with their depressive state assessed with the GDS. The 3-way ANOVA (Table 4) showed that patients' responses were not influenced by age, sex or laterality of the affected side ($p>0.24$). The Student-t test showed no significant difference between the item difficulties for items common to both the Beninese and European ABILHAND questionnaires ($p=0.72$) and ICC=0.70.

Discussion

This study presents the validation of the ABILHAND-Stroke questionnaire for the Beninese population. This self-reporting questionnaire assesses 16 bimanual activities using three-response categories. Rasch analyses demonstrated its linearity, unidimensionality, invariance and reliability. ABILHAND-Stroke is clinically relevant and well adapted to the Beninese context.

The Beninese version of the ABILHAND-Stroke questionnaire is made of 16 bimanual activities. Fourteen items are common to both Beninese and European versions and their hierarchical orders are strongly correlated. The other two items correspond to activities added by the Beninese experts during the content validity stage: "using a matchstick to light a fire" and "tying a loincloth around the waist".

The content validity is also supported by the strong correlation between the items difficulties and their bimanual requirements (Table 2). The easiest item (Fastening a snap button, -1.10 logit) belonged to group A, the moderately difficult item (Using a matchstick to light a fire, -0.24 logit) belonged to group B, and the most difficult item (Threading a needle, 2.23 logit) belonged to group C.

The ABILHAND-Stroke Benin had a lower number of items (16 items) compared to the European and Brazilian versions (23 items). Indeed, several items from the original version were discarded because they did not fit the Rasch model. There are no standard guidelines for the required number of items in a scale, yet it contributes to the reliability and targeting of the assessing tool (34). Despite this low number of items, the ABILHAND-Stroke Benin measures a difficulty range of 3.34 logits, which is targeted with the range of manual ability of the

Beninese patients. The range of measure of ABILHAND-Stroke Benin is similar to the questionnaires validated in Europe (3.90 logits) (8) and Brazil (4.11 logits) (11).

Miller et al estimated that an individual can discriminate up to 7 levels of response categories (35). However, Rasch validation often leads to reducing the number of levels. For instance, all ABILHAND questionnaires, validated for assessing manual ability among several populations and across several pathologies, included only 3 levels of responses categories. Indeed, their Rasch analyses demonstrated that patients were unable to differentiate more than 3 responses categories for manual ability (8). The number of response categories depends on the patient's discrimination capacity and the studied variable (22).

ABILHAND-Stroke Benin showed a good but relatively low PSI (0.82) compared to the European (0.90) and Brazilian (0.91) versions. However, the number of ability levels for the ABILHAND-Stroke Benin (3 levels) was slightly lower than the 4 levels of the European version and the 5 levels of the Brazilian version (11). The minimal level that a questionnaire should be able to discriminate is two (22). These lower PSI values and levels of ability might be related to the lower number of items (36).

The Rasch validation resulted in a linear scale. The results can then be submitted to arithmetic computation and parametric statistics which reduces measurement error. The linearity of the scale allowed the expression of results in percentages instead of the not well-known logit unit at using the conversion table. In clinical practice, results expressed in percentages should be more clear and meaningful for therapists and patients. Contrarily to the raw score, one per cent represents the same amount of manual ability throughout the scale. Indeed, a five-point difference on the raw score corresponds to a 25% difference at the extremities of the scale and only 9% in the middle of the scale (table 3). This is why editors recommend caution when using ordinal scales (37). Nevertheless, the Rasch method has some drawbacks. It requires specialized skills and software and a large amount of data for the questionnaire development (>100) (38).

Manual ability evaluated with the ABILHAND-Stroke Benin was well correlated with neurological impairments and manual dexterity: 36% of the variance of the manual ability score are explained by the SIAS, and 44% are explained by the affected hand BBT score (Table 4). Several other authors already underlined this relationship between the body function and structure domain and the activity domain of the ICF. Indeed, 36% of the manual ability variance can be explained by the BBT score among stroke patients (8) and 49% in children with cerebral palsy (39).

Activity limitations are related to the measures of a patient' manual ability. The manual ability was moderately-to-strongly correlated to the global activity limitations evaluated with the FIM and ACTIVLIM-stroke. In our study, 79% of the manual ability variance was explained by the ACTIVLIM-Stroke score (Table 4). Similarly, Vandervelde et al (40) showed that 58% of the manual ability variance could be explained by the global activity limitations among patients with neuromuscular disorders. One can speculate that ABILHAND and ACTIVLIM are complimentary, even though they do not assess the same variables as suggested by Vandervelde et al (40). As a matter of fact, the ACTIVLIM-stroke scale evaluates activity limitations related to the upper limb and lower limb function whereas the ABILHAND-stroke only assesses activity limitations for the former. Thus, the ABILHAND-stroke questionnaire is especially relevant in evaluations and therapies targeting upper limb functions.

In our study, hand ability measures were not correlated with patients' age or sex. These measures were also associated with neither the laterality of the affected side nor the time since stroke. These results observed along with patients' scores strengthened the results of the item location analysis in terms of differential item functioning. These results were in accordance with those reported in previous studies (8).

This study presents two limits. The coexistence of the official language (French) and multiple dialects (+/- 70) in Benin was one difficulty and represented a potential limitation for this study. To reach a large number of persons, representative of the Beninese population, the French language was used. When patients did not speak French or were unable to read French, the assessor conducted the interview in the patient's native language. On one hand, this increased the validity of the answers, but, on the other hand, this could have represented a potential bias. The targeting is suboptimal. Indeed, similarly to the Brazilians subjects (11), the mean capacity of the Beninese subjects is 1.0±1.37 logits whereas it should ideally be between -0.5 and +0.5 logit (41). But, the range of the items thresholds covers well the range of the patient's locations, and the observation of figure 2 shows that patients and items are evenly distributed.

Following the COSMIN taxonomy, several psychometric properties were demonstrated for the ABILHAND-stroke Benin questionnaire: structural, cross-cultural and convergent validity contributing to construct validity, and internal consistency contributing to reliability. However, several other psychometric properties should be further explored in a future study: test-retest reliability, measurement error and responsiveness. All patients evaluated in this study had an autonomous gait. This choice was made to avoid the impact of locomotion on manual ability, and to recruit patients who were likely to experience several manual activities in their daily environment (22). This inclusion criterion did not generate a bias favoring the recruitments of patients with greater manual ability. Indeed only 6.4% of the patients were able to easily perform all items and the patients presented with a wide range of manual dexterity (BBT range 0-75).

Conclusion

After ACTIVLIM-Stroke (19) and ABILOCO-Stroke (42), ABILHAND-Stroke Benin does increase the number of Rasch validated scales available in the African context. Together, they will help promote and standardize assessments for stroke patients' in clinical practice and improve research in this population.

References

1. Kwakkel G, Kollen BJ, Wagenaar RC. Therapy impact on functional recovery in stroke rehabilitation: a critical review of the literature. Physiotherapy 1999;13:457-70.
2. Broeks JG, Lankhorst GJ, Rumping K, Prevo AJH. The long-term outcome of arm function after stroke: Results of a follow up study. Disabil Rehabil 1999;21:357–364.
3. Pang MY, Harris JE, Eng JJ. A community-based upper-extremity group exercise program improves motor function and performance of functional activities in chronic stroke: A randomized controlled trial. Arch Phys Med Rehabil 2006;87:1-9.
4. Ashford S, Slade M, Turner-Stokes L. Conceptualisation and development of the arm activity measure (ArmA) for assessment of activity in the hemiparetic arm. Disabil Rehabil. 2013;35:1513-8.
5. World Health Organization. International Classification of Functioning, Disability and Health. Geneva: World Health Organization; 2001. Available from: http://www.who.int/classifications/icf/en/
6. Lang CE, Bland MD, Bailey RR, Schaefer SY, Birkenmeier RL. Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. J Hand Ther 2013;26:104-14.
7. Rowland TJ, Gustafsson L. Assessments of Upper Limb Ability following Stroke: a Review. British Journal of Occupational Therapy 2008;71:427-437.
8. Penta M, Tesio L, Arnould C, Zancan A, Thonnard JL The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients: Rasch-based validation and relationship to upper limb impairment. Stroke. 2001;32:1627-34.
9. Arnould C, Vandervelde L, Batcho CS, Penta M, Thonnard JL. Can manual ability be measured with a generic ABILHAND scale? A cross-sectional study conducted on six diagnostic groups. BMJ Open. 2012;2:1-9.
10. Alt Murphy M, Resteghini C, Feys P, Lamers I. An overview of systematic reviews on upper extremity outcome measures after stroke. BMC Neurol. 2015;15:1-15.
11. Basílio ML, de Faria-Fortini I, de Magalhães L, Nunes de Assumpção FS, de Carvalho AC, Teixeira-Salmela LF. Cross-cultural validity of the Brazilian version of the ABILHAND questionnaire for chronic stroke individuals, based on Rasch analysis. J Rehabil Med. 2016;48:6-13.
12. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on

measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19:539-49.

13. Gellez-Leman MC, Colle F, Bonan I, Bradai N, Yelnik A. Evaluation of the disabilities of hemiplegic patients. Ann Readapt Med Phys. 2005;48:361-8.

14. Gauthier L, Dehaut F, Joanette Y. The Bells test: a quantitative and qualitative test for visual neglect. Int J Clin Neuropsychol. 1989;11:49–54.

15. Crum RM, Anthony JC, Bassett SS, Folstein MF. Population-based norms for the mini-mental state examination by age and educational level. JAMA 1993;269:2386-91.

16. 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.

17. Mathiowetz V, Volland G, Kashman N, Weber K. Adult norms for the box and block test of manual dexterity. Am J OccupTher 1985;39:386–391.

18. Brink TL, Yesavage JA, Lum O, Heersema PH, Adey M, Rose TL. Screening tests for geriatric depression. Clin Gerontol. 1982;1:37–43.

19. Batcho CS, Tennant A, Thonnard JL. ACTIVLIM-Stroke: across cultural Rasch-built scale of activity limitations in patients with stroke. Stroke. 2012;43:815-23.

20.Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation. J Rehabil Med 2003;35:105–115.

21. Rasch G**.** Probabilistic models for some intelligence and attainment tests. Chicago; University of Chicago; 1960.

22. Penta M., Arnould C., Decruynaere C., Développer et interprèter une échelle de mesure: Applications du modèle de Rasch. Brussels : Edition Mardaga; 2005.

23. Andrich D. Category ordering and their utility. Rasch Measurement Transactions. 1996;9:464-5.

24. Zhu W., Updyke WF., Lewandowski C. Post-hoc Rasch analysis of optimal categorization of an ordered-response scale. J Outcome Meas 1997;1:286-304.

25. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3:205-31.

26. 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.

27. Lo C, Liang WM, Hang LW, Wu TC, Chang YJ, Chang CH. A psychometric assessment of the St. George's respiratory questionnaire in patients with COPD using Rasch model analysis. Health Qual Life Outcomes. 2015;20:13:1-16.

28. Smith RM. IPARM : Item and Person analysis with Rasch model. Chicago: MESA Press; 1991.
29. Andrich D, editor. Rasch analysis for measurement. London : Sage Publications Ltd ; 1988.
30. Holland PH, Wainer H. Differential Item Functioning. Hillsdale, NJ : Lawrence Erlbaum 1993.
31. Wright DB, Stone M.H. Best Test Design. Chicago : MESA Press 1979.
32. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. 2nd ed. Mahwah: Lawrence Erlbaum Associates; 2007.
33. Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil. 2000;81(12 Suppl 2):S15-20.
34. Hinkin, T. R., Tracey, J. B., & Enz, C. A. Scale construction: Developing reliable and valid measurement instruments. Journal of Hospitality & Tourism Research 1997;21:100-120.
35. Miller GA. The magical number seven, plus or minus two : some limits to our capacity for processing information. Psychological Review. 1956;63:81-97.
36. Frisbie DA. Reliability of scores from teacher-made tests. Educational Measurement: Issues and Pratical, National council on Measurement in Education. 1988;7:25-35.
37. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: time to end malpractice? J Rehabil Med. 2012;44:97-8.
38. Institute for Objective Measurement (IOM). The Rasch Model as a Construct Validation Tool. Available from: http://www.rasch.org/rmt/rmt221a.htm
39. Arnould C, Penta M, Thonnard JL. Hand impairments and their relationship with manual ability in children with cerebral palsy. J Rehabil Med. 2007;39:708-14.
40. Vandervelde L, Van den Bergh PY, Penta M, Thonnard JL. Validation of the ABILHAND questionnaire to measure manual ability in children and adults with neuromuscular disorders. J Neurol Neurosurg Psychiatry 2010;81:506-12.
41. Linacre J.M. Sample size and item calibration stability. Rasch Measurement Transactions, 7, 28. 1994.
42. Sogbossi ES, Thonnard JL, Batcho CS. Assessing locomotion ability in West African stroke patients: validation of ABILOCO-Benin scale. Arch Phys Med Rehabil. 2014;95:1470-6.

Table 1: Patient characteristics.

Age in years: median[Range]	55 [28-79]
Sex	
Male	149(66.8%)
Female	74(33.2%)
Stroke type	
Ischemic	79(35.4%)
Hemorrhagic	33(14.8%)
Undefined (brain CT-Scan not available)	111(49.8%)
Paretic side	
Dominant	108(48.4%)
Non dominant	115(51.6%)
Time since stroke	
Delay in month: Median [Range]	27[6-99]
≤ 2 years	99(44.4%)
>2 years	124(55.6%)
FIM-Mo: median [Range]	87[30-87]
BBT: median [Range]	32[0-75]
SIAS: median [Range]	65[19-76]
GDS (n=179)	
Median [range]	9[0-27]
< 10	87(48.6%)
≥10	92(51.4%)

FIM-Mo: Motor subscale of the Functional Independence Measure. BBT: Box and Block Test. SIAS: Stroke Impairment Assessment Set. GDS: Geriatric Depression Scale

Table 2: Item difficulty location and fit index for the 16 items of the questionnaire.

Items	Difficulty (logit)	SE (logit)	Residuals	Chi-square	DF	P value*	Bimanual implication
Fastening a snap	-1.10	0.19	-0.04	5.48	2	0.06	A
Opening a jar	-1.04	0.20	-0.13	0.99	2	0.61	B
Spreading toothpaste	-1.02	0.17	-1.02	1.70	2	0.43	A
Closing the zipper on	-0.91	0.20	0.55	1.17	2	0.55	A
Opening a chocolate	-0.89	0.2	0.93	3.88	2	0.14	B
Buttering a slice of	-0.86	0.19	-0.88	2.58	2	0.28	B
Opening a letter	-0.74	0.23	-0.17	3.06	2	0.22	B
Using a matchstick to	-0.24	0.16	-0.43	0.73	2	0.69	B
Opening a bottle	-0.22	0.17	-0.38	3.72	2	0.15	B
Tying a loincloth	0.20	0.19	0.09	0.54	2	0.76	C
Buttoning trousers	0.39	0.15	-1.13	2.13	2	0.34	B
Cutting meat with a	0.66	0.18	0.06	0.55	2	0.75	C
Sharpening a knife	0.83	0.17	0.82	3.45	2	0.17	C
Hammering a nail	1.33	0.16	-0.97	5.69	2	0.05	C
Clipping nails	1.38	0.16	0.27	2.86	2	0.24	C
Threading a needle	2.23	0.17	0.11	3.78	2	0.15	C

*p-value after Bonferroni adjustment = 0.002. DF=Degrees of Freedom.

Table 3: Ordinal-to-interval score transformation table

Ordinal score	Interval score in logit	Interval score in percentage
0	-4.37	0
1	-3.54	9
2	-2.97	16
3	-2.57	20
4	-2.26	23
5	-1.99	26
6	-1.77	29
7	-1.56	31
8	-1.37	33
9	-1.19	35
10	-1.02	37
11	-0.86	39
12	-0.69	41
13	-0.54	43
14	-0.39	44
15	-0.23	46
16	-0.08	48
17	0.08	49
18	0.23	51
19	0.4	53
20	0.57	55
21	0.75	57
22	0.93	59
23	1.13	61
24	1.33	63
25	1.55	66
26	1.79	68
27	2.07	71
28	2.37	75
29	2.72	79
30	3.17	84
31	3.79	91
32	4.64	100

In clinical practice, if patient answered all questions, the ordinal score can be easily transformed in interval score, expressed in logit or in percentage. If the patient did not answer all questions, then a software or a website (for instance http://www.rehab-scales.org) is required to transform the ordinal score into interval score.

Table 4: relationship between the manual ability score of patients and population characteristics and functional assessment.

	Statistics	p-value
Age	F=0.10	0.74
Sex	F=1.35	0.24
Affected side	F=0.14	0.71
BBT	rho=0.66	<0.001
SIAS	rho=0.60	<0.001
GDS	rho=-0.24	<0.001
FIM-Mo	rho=0.56	<0.001
ACTIVLIM-Stroke	r=0.89	<0.001

BBT: Box and Block Test. SIAS: Stroke Impairment Assessment Set. GDS: Geriatric Depression Scale. FIM-Mo: Motor subscale of the Functional Independence Measure. F = three-way (age, sex and affected side) analysis of variance assessing the effect of age, sex and affected side on the person location.

The correlation coefficients (r = Pearson; rho = Spearman) present the relation between the patients' manual ability and their gross manual ability (Box and Block test), neurological impairments (SIAS), depression state (GDS), functional independence (FIM-Mo and Activlim).

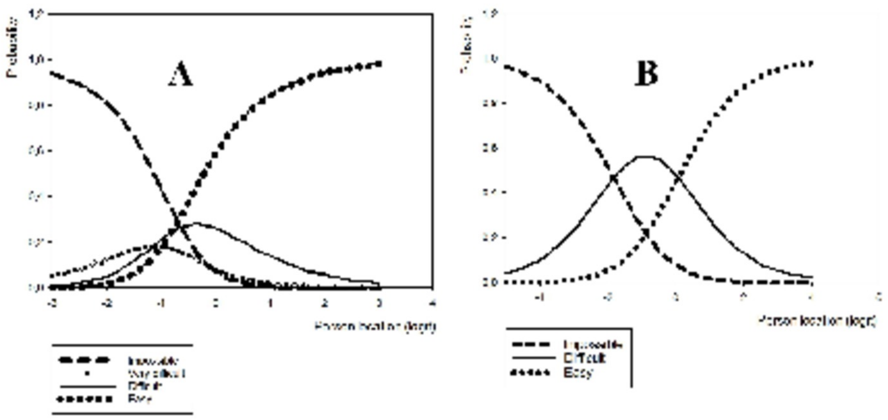

Figure 1: curve of the probabilities of response categories. A = curve with four categories of responses showing disordered thresholds for the response categories. B= curve with three categories of responses, no disordered thresholds.

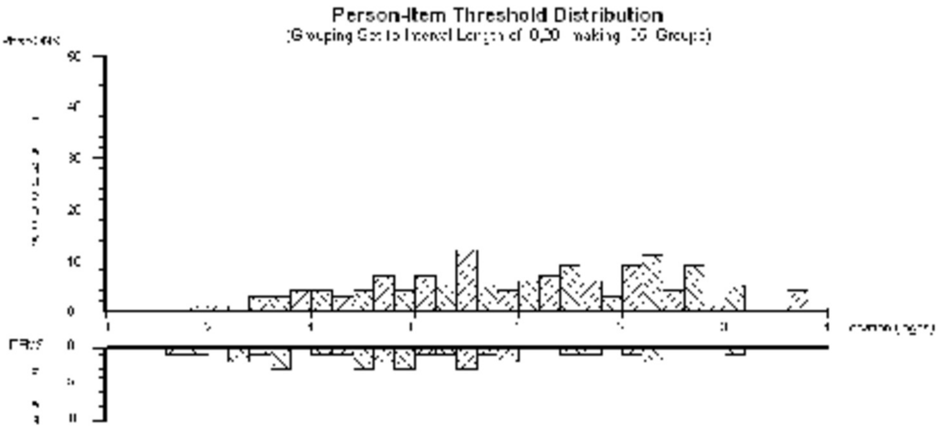

Figure 2: Person item distribution. The upper graph shows the frequency of person distribution according to their ability; the graph below shows the frequency of item distribution according to their difficulty. The x-axis shows items and persons location on the same scale; the y-axis we shows distribution frequencies.

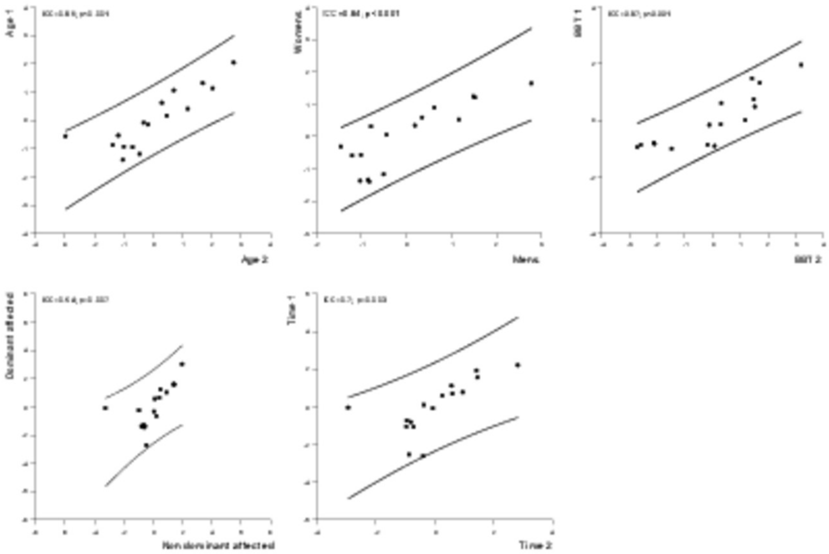

Figure 3: ABILHAND-Stroke questionnaire invariance via seven subgroups: age (age1: ≤ 55 years; age2: >55 years), sex, time since stroke (time1: ≤ 2years; time 2: > 2years), laterality of the affected side, motor impairment score with the SIAS (SIAS1: ≤65; SIAS2: >65), manual dexterity score of the affected side evaluated with the BBT (BBT1: ≤32; BBT2: >32); depression level based on the GDS (GDS1: <10; GDS2: ≥10) scores. Lines indicate a confidence interval of 95%. Points outside of the two lines would indicate items presenting with differential item functioning. No item of the questionnaire presents with differential functioning.