J Rehabil Med 2016; 48: 6–13

ORIGINAL REPORT

CROSS-CULTURAL VALIDITY OF THE BRAZILIAN VERSION OF THE ABILHAND QUESTIONNAIRE FOR CHRONIC STROKE INDIVIDUALS, BASED ON RASCH ANALYSIS

Marluce Lopes Basílio, MSc[1], Iza de Faria-Fortini, MSc[2], Lívia de Castro Magalhães, PhD[2], Fernanda Sabine Nunes de Assumpção, MSc[1], Augusto Cesinando de Carvalho, PhD[3] and Luci Fuscaldi Teixeira-Salmela, PhD[1]

From the [1]Department of Physical Therapy, [2]Department of Occupational Therapy, Universidade Federal de Minas Gerais, Belo Horizonte and [3]Department of Physical Therapy, Universidade Estadual Paulista, Presidente Prudente, Brazil

Objective: **To evaluate the cross-cultural validity of the Brazilian version of the ABILHAND for stroke individuals.**
Subjects: **A total of 107 community-dwelling chronic stroke survivors; mean age 58 years.**
Methods: **Cross-cultural adaptation of the ABILHAND followed standardized procedures. Measurement properties of the adapted version were analysed using Rasch analysis. Cross-cultural validity was based on cultural invariance analyses.**
Results: **The ABILHAND-Brazil demonstrated satisfactory performance as a rating scale. Only one item exhibited misfit to the Rasch model expectations. Principal component analysis of the residuals showed that the manual ability of the individuals encompassed different contents related to the degree of the paretic upper limb involvement in performing manual activities. Some minor local dependency was identified in 2 pairs of items (residual correlations > 0.3). Furthermore, the adapted version exhibited high levels of reliability, no floor effects, and minimal ceiling effect. Analyses of cultural invariance showed that the ABILHAND-Original and ABILHAND-Brazil calibrations can be used interchangeably.**
Conclusion: **The ABILHAND specific for stroke individuals demonstrated satisfactory measurement properties for use within both clinical and research contexts in Brazil, and cross-cultural validity for use in international/multicentric studies between Brazil, Belgium, and Italy.**

Key words: stroke; activities of daily living; upper extremity; questionnaires; psychometrics.

J Rehabil Med 2016; 48: 6–13

Correspondence address: Luci Fuscaldi Teixeira-Salmela, Department of Physical Therapy, Universidade Federal de Minas Gerais, Avenida Antônio Carlos, 6627, Campus Pampulha, BR-31270-901 Belo Horizonte, MG, Brazil. E-mail: lfts@ufmg.br

Accepted Oct 14, 2015; Epub ahead of print Dec 11, 2015

INTRODUCTION

Stroke is the main cause of death and disability in Brazil (1), with an age-standardized incidence between 216 and 252 per 100,000 persons/year (2). A Brazilian prospective population-based study reported that only 43% of patients were independent in activities of daily living (ADL) one year after stroke (3).

It is well-known that limitations in ADL are strongly associated with upper-limb (UL) impairments (4, 5), which are common after stroke (6, 7). However, these relationships are not always univocal and predictable (8), since activity limitations depend on complex interactions between motor function and contextual factors (9, 10). Thus, these limitations should be independently measured using specific tests, in order to explore their associations with other constructs (8).

The ABILHAND is a questionnaire for the assessment of manual ability, defined as the ability to manage daily activities that require the use of the UL, regardless of the strategies involved (9). The version for stroke individuals has shown construct validity (9), adequate test-retest reliability (11), and clinical utility (12). The ABILHAND was constructed according to the Rasch measurement model, which allows for the conversion of ordinal scores into linear measures (9). The main advantage of this is that linear measures are expressed on scales with equal units, allowing correct inferences from direct inter- or intra-individual comparisons (13, 14).

Many studies have addressed the cross-cultural validity of outcomes (15–17). The ABILHAND has great potential to be used within cross-cultural contexts, since its cross-cultural validity has already been examined between Belgian and Italian samples with good results (9). However, its cross-cultural validity should also be examined with samples from other countries, so that it can be used consistently in international studies.

Great emphasis has been placed on Rasch analysis for the evaluation of cross-cultural validity (10, 15–18), since it allows the investigation of cultural invariance, i.e. if a scale works in the same way regardless of the country in which it is applied (15). The procedure is based on examination of differential item functioning (DIF), i.e. items that perform differently from one group to another (19, 20). Cultural invariance of a scale is required to ensure that equivalent scores will represent equivalent levels of the measured construct across different populations (16).

 doi: 10.2340/16501977-2044
 ISSN 1650-1977

The aim of this study was, therefore, to evaluate the cross-cultural validity of the ABILHAND specific for stroke individuals. First, the process of cross-cultural adaptation (21) of the ABILHAND was conducted to enable its application in Brazil. Rasch analysis was performed to investigate whether the adapted version had adequate properties for the intended application (appropriateness of the rating scale, construct validity, reliability, sample targeting, and local independence) (20). This analysis also allows for the evaluation of cross-cultural validity (16).

METHODS

Participants

Community-dwelling people with stroke, living in the city of Belo Horizonte, Brazil, were recruited from the general community through advertisements, from screening out-patient clinics in public rehabilitation services, and from lists of previous research projects.

Inclusion criteria were: ≥20 years of age; mean time since onset of unilateral stroke at least 6 months; no cognitive deficits, as determined by the Mini-Mental State Examination cut-off scores (22); clinical signs of hemiparesis, i.e. increased tonus of the elbow flexor muscles, determined by scores different from zero on the Modified Ashworth Scale (23) and/or weakness of the shoulder flexors, elbow flexors/extensors, wrist extensors, and finger flexors, determined by differences above 10% between the paretic and non-paretic UL (24, 25). Exclusion criteria were: individuals unable to express themselves verbally, those with uncorrected visual deficits, bilateral hemiparesis, and other disabling musculoskeletal or neurological conditions.

This study was approved by the ethical review board of the Universidade Federal de Minas Gerais, Brazil, and all participants provided written consent prior to data collection.

ABILHAND

The ABILHAND, specific for stroke individuals, contains 23 bimanual activities, rated as: 0 = Impossible, 1 = Difficult, or 2 = Easy (9). It is administered by interviews, during which the individuals are asked to estimate their ability to perform the activities without help, irrespective of the limb(s) actually used to perform them and the strategy used (26). Activities not attempted during the previous 3 months are not scored and are entered as missing responses (26). The responses should be submitted to Rasch analysis, which, from the ordinal scores, calibrates the ability of the individuals and the difficulty of the items in a linear continuum (scale) divided into equal units (logits). The manual ability measure is equivalent to the individual's position along this scale (9, 26).

Procedures

Cross-cultural adaptation. Cross-cultural adaptation of the ABILHAND followed recommended procedures (21, 27) and was carried-out in 5 stages. First, the ABILHAND was translated from English to Brazilian Portuguese, independently, by 2 bilingual translators, whose native language was Portuguese. Secondly, a synthesis of the translations was produced, followed by back-translation, which was carried out independently by 2 other bilingual translators, whose native language was English. Neither translator had access to the original version, or was informed about the concepts of the questionnaire. An expert committee, composed of 3 physical therapists, 1 occupational therapist, 1 translator, and 1 back-translator, then, consolidated all versions of the questionnaire and developed its pre-final version. Finally, the pre-final version was administered to 10 individuals with chronic stroke, who responded to the questionnaire and were asked to interpret each item. As there was no problem regarding the wording and clarity of the items, the final version, ABILHAND-Brazil, was established.

Application of ABILHAND-Brazil. Initially, all participants were physically screened to verify the eligibility criteria. Demographic and clinical information, such as the time since the onset of stroke, paretic side, UL dominance previous to stroke, and UL motor recovery, which was assessed using the Fugl-Meyer assessment (FMA) scale – upper limb section (28, 29), were obtained for characterization purposes. The ABILHAND-Brazil was then individually applied, following standardized procedures (26). All data were collected by well-trained physical therapists in a research laboratory setting.

Data analyses

Rasch analysis. The process and concepts related to Rasch analysis are discussed in dedicated textbooks (14, 30) and articles (20, 31). Rasch analysis was performed using the WINSTEPS software, version 3.81.0. The rating scale model was used, considering that the 3-point rating scale of the ABILHAND should work similarly for all items (20).

Analyses were carried out, as follows:

Rating scale analysis. The appropriateness of the 3-point rating scale of the ABILHAND-Brazil was evaluated according to the following criteria: at least 10 responses per category, monotonically increased in both mean measures and Andrich thresholds (step calibrations) across categories, and outfit mean-square (MnSq) values lower than 2 (14). Correct category discrimination is necessary to provide true information regarding the person's location on the variable (32).

Construct validity. Construct validity was verified by evaluation of the unidimensionality of the ABILHAND-Brazil, by means of fit statistics and principal component analysis (PCA) of the standardized residuals (14). To examine how well the items fitted the model expectations, goodness-of-fit statistics were considered in 2 formats, infit and outfit (MnSq) in combination with standardized Z values (Zstd) (14). The critical values for a Type 1 error rate of 5% were calculated using the following formulae, which considered the influence of the size of the sample: MnSq (infit) = $1+ 2/\sqrt{x}$; MnSq (outfit) = $1+6/\sqrt{x}$, where "x" is the sample size (33). Items with MnSq > critical values in combination with $Z>2$ indicated that the responses were erratic, i.e. misfit (14). When more than 5% of the total number of the items are erratic, this is a great threat to the construct validity, because it indicates that the items do not combine to measure a unidimensional construct (34). In addition, the same fit statistics and criteria were used for the examination of person fit. This analysis is also important, because individuals with erratic responses may affect the item fit (20).

For the PCA, to characterize unidimensionality, the criteria used were that the principal component (large dimension) should explain at least 50% of the total residual variance and, after removal of this component, a second large dimension should explain less than 5% of the remaining variance (19) or have an eigenvalue lower than 2 (30, 35). If a second dimension is identified, it is necessary to evaluate whether its size and nature would justify separate analysis for each dimension (14).

Local dependence. The PCA also allows for the evaluation of local dependence between items. After removal of the principal component, the existence of local dependency was determined by examination of the residual correlation matrix. Correlations above 0.3 were considered indicative of dependence between the set of items (32, 36), either because they duplicated some feature of each other or because they both incorporated some other shared dimension (30).These items could inflate the classic reliability and parameter estimation in Rasch analysis (20). Where high residual correlations were detected, these items were combined into "testlets" and their effects were re-analysed (30, 36).

Reliability. Both person and item separation coefficients were used to estimate the number of strata within the range of the observed persons' abilities and item difficulties (30). To calculate the number of strata, the following formula was employed: number of strata = $(4G+1)/3$, where G is the separation coefficient (30). It was expected that the individuals were stratified into at least 2 strata (low and high abilities), which would imply a person reliability index > 0.80, and that the items were stratified into at least 3 levels of difficulty (low, medium and high), which would result in an item reliability index > 0.90 (30). If these criteria were not satisfied, more data should be collected, to reduce the error or imprecision of the estimates (14).

Item-person map. This map is a visual representation of the manual ability scale, in which both items and individuals are displayed along the same linear continuum (19). This allows for the investigation of whether the ABILHAND items were appropriate for the ability levels of the sample, ceiling/floor effects, and gaps (i.e. few or no items in certain ability level) (14, 19). On the map, the individuals were identified by gender, age, previous dominance of the paretic UL, and UL motor recovery, to determine whether these variables affected their manual abilities.

Cross-cultural validity. The cross-cultural validity of the ABILHAND-Brazil was based on cultural invariance analyses of both the estimates of item difficulty (DIF analysis) and the estimates of persons' ability (14). A DIF plot was used to compare the item calibration of ABILHAND-Original (Belgian and Italian samples) (9) with that of the ABILHAND-Brazil. If an item's difficulty estimate varies across the samples by more than the modelled error (i.e. the item location fell outside the 95% confidence intervals), this is the *prima facie* evidence for lack of measurement invariance across countries (14). To examine the invariance of person's ability estimates, the responses of the sample of this study were then anchored with the item calibration of the ABILHAND-Original (14). The manual ability estimates obtained by this anchoring were, then, compared with the estimates of the same sample obtained with the calibration of ABILHAND-Brazil, by means of a scatter-plot (14).

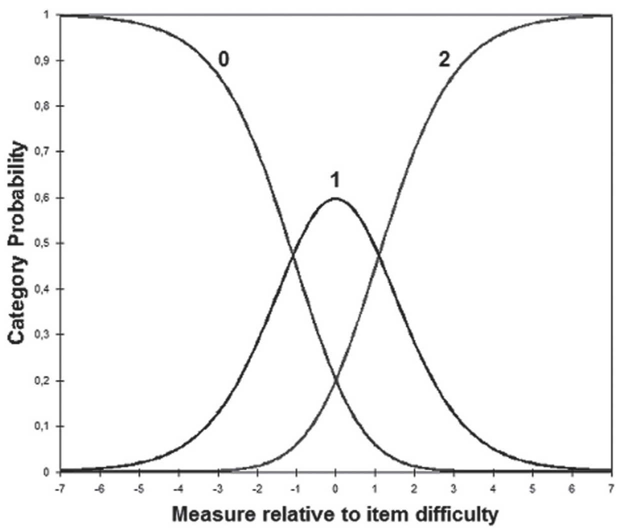

Fig. 1. Probability curves for the ABILHAND-Brazil rating scale. This graph illustrates the probability of responding to any particular category, given the difference in estimates between any person ability and any item difficulty. Each category (0, 1, or 2) had a distinct peak in the probability curve graph, illustrating that each was the most probable response category for some portion of the manual ability. The threshold estimates correspond to the intersection of the rating scale categories.

RESULTS

Participants' characteristics

From a list of 485 individuals, 121 agreed to participate and were physically screened. Fourteen individuals did not meet the inclusion criteria; thus, 107 chronic stroke individuals participated in the present study; 59% men, mean age 58 years (standard deviation; SD 12), and mean time since stroke 64 (SD 64) months (range 6–380 months). Participants' mean FMA (UL section) scores were 44 (SD 19), with 5% of individuals classified as normal, 47% as having mild impairments, 22% moderate impairments, and 26% severe impairments. In 55% of the subjects, the paretic UL was the dominant limb before stroke.

Rasch analysis

Rating scale analysis. Rasch analysis showed sufficient frequency counts and category fit at all 3 levels. The mean value of measures across the categories (0–2) increased monotonically and none of these exhibited disordered step calibrations (Fig.1). This indicated that the rating scale structure was adequate.

Construct validity. The critical infit and outfit MnSq values were > 1.2 and > 1.6, respectively. As shown in Table I, only item 12 (Tearing open a pack of chips) exhibited marginal misfit (Infit MnSq = 1.45, Zstd = 3.2; Outfit MnSq = 1.41, Zstd = 2.2). Further investigation showed that for 8 of the 10 individuals, who showed residuals in this item, had their paretic UL as dominant before the stroke. Out of the 107 individuals, 7 (6.5%) exhibited misfit, which is close to the recommended 5% value. Thus, none of the individuals were removed from the subsequent analyses. The PCA revealed that 51.9% of the total variance was explained by the principal component, but the eigenvalue of 3.0 was related to a second component, which explained 6.3% of the remaining variance. Although these results suggested the existence of a second dimension, the contrast plot showed that the items at the top and bottom were not different enough to be considered 2 dimensions. The top items, "Peeling onions", "Peeling potatoes with a knife" and "Sharpening a pencil", are tasks that require digital activity from the paretic UL, whereas the bottom items, "Fastening a snap", "Pulling up the zipper of trousers" and "Buttoning up trousers", require stabilization of the paretic UL or involve unimanual movement sequences (9). These contents are strands within the main dimension of manual ability.

Local dependency. The assessment of local dependency identified the following 2 pairs of items with correlations > 0.3: "Peeling potatoes with a knife" and "Peeling onions" (r = 0.50), "Buttoning up trousers" and "Pulling up the zipper of trousers" (r = 0.43). However, the results of the re-analysis after combining these locally dependent items into 2 testlets showed no meaningful differences, compared with the analyses of the original single items. There was a minor decrease in reliability estimate (person reliability index = 0.90, Cronbach's alpha = 0.91), compared with the original estimates (person reliability index = 0.91, Cronbach's alpha = 0.92). The item 12 "Tearing open a pack of chips" still did not fit the Rasch model (Infit MnSq = 1.41, Zstd = 2.9; Outfit MnSq = 1.36, Zstd = 2.0) and the improvement in the PCA dimensionality parameter was small (eigenvalue 2.6 for the second component).

Table I. *Calibration and item fit statistics of the ABILHAND-Brazil*

			Infit		Outfit	
Items	Difficulty (logits)	SE (logits)	MnSq	Zstd	MnSq	Zstd
4. Cutting one's nails	2.07	0.19	0.95	–0.3	0.85	–09
2. Threading a needle	1.98	0.19	1.21	1.4	1.29	1.6
6. Filing one's nails	1.48	0.19	1.00	0.0	1.04	0.3
5. Wrapping up gifts	1.05	0.19	0.75	–2.0	0.88	–0.8
3. Peeling potatoes with a knife	1.02	0.18	0.81	–1.6	0.72	–2.0
1. Hammering a nail	0.99	0.18	1.17	1.3	1.15	1.0
7. Cutting meat	0.88	0.18	0.91	–0.7	0.83	–1.2
9. Shelling hazelnuts	0.59	0.20	1.22	1.5	1.09	0.5
14. Sharpening a pencil	0.49	0.19	1.01	0.1	0.90	–0.6
12. Tearing open a pack of chips	**0.27**	**0.18**	**1.45**	**3.2**	**1.41**	**2.2**
8. Peeling onions	0.25	0.18	1.01	0.1	0.90	–0.6
10. Opening a screw-topped jar	–0.19	0.18	1.22	1.6	1.33	1.5
13. Buttoning up a shirt	–0.33	0.18	0.78	–1.7	0.90	–0.4
11. Fastening the zipper of a jacket	–0.57	0.19	1.06	0.5	0.97	–0.1
15. Spreading butter on a slice of bread	–0.59	0.19	0.78	–1.7	0.64	–1.7
16. Fastening a snap (e.g. jacket, bag)	–0.64	0.19	1.14	1.0	1.01	0.1
17. Buttoning up trousers	–0.69	0.19	0.98	–0.1	0.82	–0.7
22. Unwrapping a chocolate bar	–0.71	0.19	0.94	–0.4	0.90	–0.3
18. Taking the cap of a bottle	–0.76	0.19	1.00	0.0	1.19	0.8
19. Opening mail	–1.14	0.20	0.85	–1.0	0.94	–0.1
21. Pulling up the zipper of trousers	–1.57	0.22	0.77	–1.4	0.79	–0.5
23. Washing one's hands	–1.82	0.23	1.02	0.2	0.76	–0.5
20. Squeezing toothpaste on a toothbrush	–2.04	0.24	0.97	–0.1	0.71	–0.5

Misfitting item is shown in bold.
SE: standard error; MnSq: mean square; Zstd: standardized Z value.

Reliability. Item separation analysis indicated that the items were distributed into 7.9 levels of difficulty, leading to a reliability index of 0.97. Person separation analysis indicated that 4.5 different levels of ability could be distinguished in this sample, with a person reliability index of 0.91.

Item-person map. The item-person map (Fig. 2) showed that most of the items fell in the middle third of the continuum, where the ability of most individuals were also located. However, it can be seen that some individuals are at the top without aligned items. In fact, the mean ability of the sample is only 1.0 (SD 1.64) logit above the average difficulty of the items. In addition, there was no floor effect and the ceiling effect was lower than 1%, since only 1 subject scored all items as easy. Regarding the subject's characteristics, only motor recovery showed irregular distribution along the continuum, where individuals with mild impairments were concentrated at the top and those with severe impairments at the bottom (Fig. 2).

Cross-cultural validity. Fig. 3 (A) shows the results of the DIF analysis, in which the items 1, 11, 14 and 22 (refer to Table I for item descriptions) showed DIF across the samples of Italy/Belgium and Brazil. However, the results of subsequent analysis showed invariance of person estimates, because the measures, in logits, obtained with the ABILHAND-Brazil and ABILHAND-Original calibrations, were almost identical (Fig. 3B). Although there were differences in item calibrations across the samples from different countries, they did not affect the estimates of manual ability.

DISCUSSION

The present study evaluated the cross-cultural validity of the ABILHAND for individuals with stroke. The ABILHAND-Brazil, including the manual and the questionnaire in 10 random orders, is available free of charge from www.rehab-scales.org (26).

Rasch analysis showed that the rating scale (Impossible, Difficult, or Easy) functioned well to create an interpretable measure of manual ability. All categories were used often, indicating that none was unnecessary or redundant (14). The average measures and the Andrich thresholds across the categories increased monotonically, which meant that the changes from the lowest to the highest category were followed by increases in manual ability (14). Finally, all categories showed adequate fit, i.e. all contributed with important information to discriminate the manual ability levels of the participants (14).

Goodness-of-fit analysis showed that only the item "Tearing open a pack of chips" was identified as misfitting and showed a high number of erratic responses by the individuals, who had their paretic UL as dominant before the stroke. One hypothesis that could explain this result is random guessing. In fact, self-reported measures are prone to over- or under-estimated responses (9), which may occur in items that are too difficult for the respondent (14). It is likely that this item was very difficult for these individuals. However, even with 1 misfitting item (4.3%), the ABILHAND- Brazil met the limit of 5% of erratic items, indicating that the items contributed to the definition of a unidimensional scale. It is noteworthy

```
MEASURE                | BOTTOM P=50%  | MEASURE       | TOP P=50%     MEASURE
  <more> ----- PERSON -+- ITEM        -+- ITEM        -+- ITEM          <rare>
    5                N +               +               +                  5
                       |               |               |
                    NN |               |               |
                     M |               |               |
                       |               |               |
                       |               |               |
    4             LLLL +               +               +                  4
                       |               |               |
                       |               |               |
                     N |               |               |
                     L |               |               |
                    NL |               |               | 2.4.
    3               LM +               +               +                  3
                   LLL |               |               |
                       |               |               | 6.
                   MLL |               |               |
                   LLM |               |               |
                 SLLLM |               |               | 1.3.5.
    2                  +               + 2.4.          + 7.               2
                  MLML |               |               | 9.
                 LLMML |               |               | 14.
                 LLLNL |               | 6.            | 12.
                   LLL |               |               | 8.
                   LLM |               |               |
    1             LLLM +               + 1.3.5.        + 10.              1
                     S | 2.4.          | 7.            | 13.
                 LSMLS |               | 9.            | 11.
                 LSLLM |               | 14.           | 15.16.17.22.
                    MS | 6.            | 12.           | 18.
                 LSMLL |               | 8.            |
    0             MMSM +               +               + 19.              0
                LMLSMS | 1.3.5.        | 10.           |
                 SSSSL | 7.            | 13.           | 21.
               SSLLSLS | 9.            | 11.           |
                   MSS | 14.           | 15.16.17.22.  | 23.
                   MSS | 12.           | 18.           | 20.
   -1                  + 8.            +               +                 -1
                    SS |               | 19.           |
                    SS | 10.           |               |
                    LM | 13.           | 21.           |
                       | 11.           |               |
                       | 15.16.17.22.  | 23.           |
   -2                  + 18.           + 20.           +                 -2
                    MS |               |               |
                     S | 19.           |               |
                       |               |               |
                       | 21.           |               |
                       |               |               |
   -3                  + 23.           +               +                 -3
                       | 20.           |               |
                       |               |               |
                       |               |               |
                       |               |               |
                       |               |               |
   -4                  +               +               +                 -4
  <less> ----- PERSON -+- ITEM        -+- ITEM        -+- ITEM          <frequent>
```

Fig. 2. Item-person map of the ABILHAND-Brazil. The left-hand column (*first column*) locates the person ability measures along the continuum of manual ability. All 107 individuals are represented as their upper limb motor recovery levels, assessed by the Fugl-Meyer scale (N=normal, L=light (mild), M=moderate, S=severe). On the right columns, each item is shown 3 times. In the centre column (*third column*), each item is placed at its mean calibration. Step calibrations 0–1 and 1–2 for the rating scale are presented in the second and fourth columns, respectively.

that, when measuring a complex attribute related to human behaviour, such as manual ability (9), one cannot expect that all items and persons fit perfectly in a strictly mathematical expression, such as the Rasch model (14).

The PCA suggested the existence of a second dimension of the ABILHAND-Brazil. However, the contrast plot showed that the items were separated according to the involvement of the paretic UL, when performing the activities. This in itself did not threaten the unidimensionality of the scale, but showed that manual activities include several levels of bimanual involvement. Furthermore, corroborating the findings of the study that validated the ABILHAND for stroke individuals (9), in the present study the most difficult items were those that required high involvement of the paretic UL, whereas the easiest ones were those that only required stabilization of the paretic UL or involved unimanual movement sequences. It is expected that a questionnaire should include both easy and difficult items for the assessment of individuals with various levels of ability, which provides higher sensitivity and precision to the measure (37). Thus, also considering the adequate results of the goodness-of-fit analysis, these contents did not harm the theoretical meaning or use of the measure, but provided

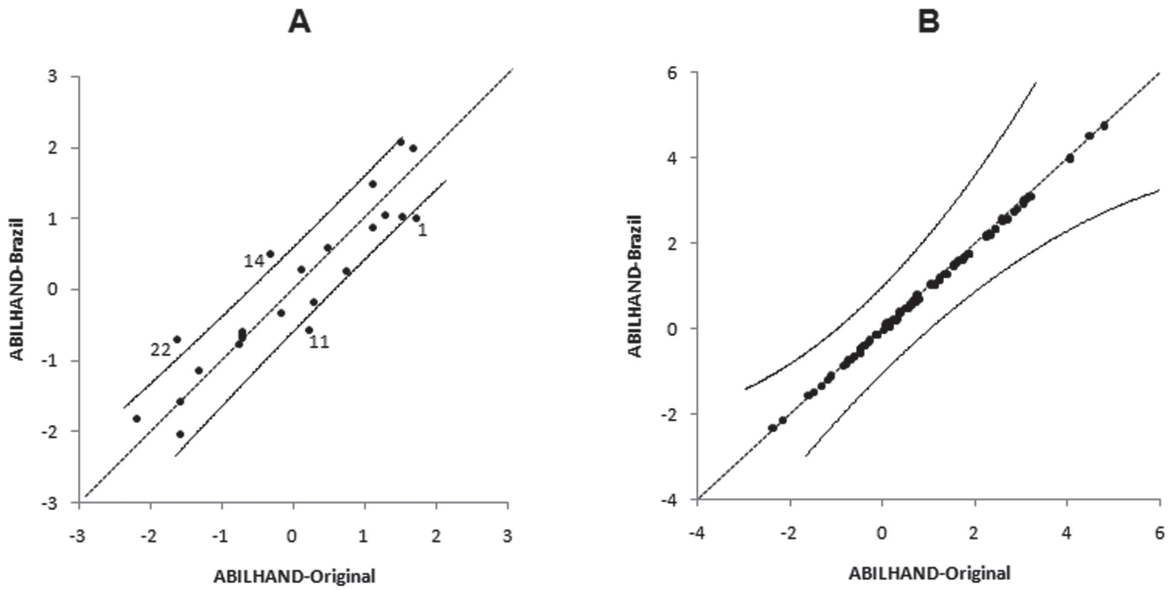

Fig 3. (A) Differential item functioning (DIF), by comparing the calibrations of the ABILHAND-Brazil and ABILHAND-Original. The items that showed DIF are labelled; (B) Comparison between the manual ability measures of the Brazilian sample obtained with the ABILHAND-Brazil and the anchored ABILHAND-Original calibrations. For the 2 graphs, the x- and y-axes show the estimates, in logits. The equality model (x = y) is represented by the identity line (*dashed line*). The control lines (*solid lines*), constructed from the standard error values of each pair of estimates, determine the 95% confidence band around the identity line. The points outside of the control lines show differences between the pair of the estimates.

a clinical sense in the item hierarchy. Therefore, the results of this analysis appeared to be more due to the existence of 2 manual ability sub-dimensions, related to the degree of the paretic UL involvement in performing the activities, than to the existence of a second dimension. Separate measurement of these sub-dimensions would not be feasible, because it could result in measures for the restricted use of specific levels of manual ability.

Some degree of local dependency between items was expected, due to their content. The 2 pairs of locally dependent items may have contributed to the PCA results, because the items "Peeling potatoes with a knife" and "Peeling onions" are at the top of the contrast plot. On the other hand, the items "Buttoning up trousers" and "Pulling up the zipper of trousers" are at the bottom. However, re-analysis, after combining these pairs of items into testlets, did not significantly improve the dimensionality of the scale. Furthermore, these items did not result in artificial inflation of the reliability estimates, indicating no serious impact on the discriminative capacity of the measure for individual application (38). Therefore, there is no stronger rationale to recommend revision or even exclusion of these items.

The separation coefficients of the persons and items determined the levels of ability/difficulty sufficiently spread along the manual ability continuum, without large agglomerations. The scale was able to distinguish nearly 5 levels of manual ability and nearly 8 levels of items difficulty. These results led to high indices of reliability of both persons and items, indicating that the questionnaire would be reproducible over time, and able to provide reliable measures (10, 14).

The item-person map showed some individuals on the top of the continuum, without the presence of aligned items. However, the high levels of reliability, the lack of floor effect, and a minimal ceiling effect suggested that the ABILHAND-Brazil is appropriate to measure individuals with varied abilities. Caution should be taken, when applying this questionnaire to individuals with high levels of manual ability, since the scale does not have enough difficult items to measure their levels of ability with precision. This suggests that the ABILHAND has greater potential to measure more severely disabled individuals. Regarding the participants' characteristics, only motor recovery appeared to have influenced manual ability, since individuals with mild impairments were concentrated at the top of the continuum and those with severe impairments at the bottom. These observations were coherent with the results of Penta et al. (9), which showed that motor deficits were one of the most influential determinants of the ABILHAND scores for stroke individuals. In addition, they also did not find any relationships between manual ability and age, sex, and previous dominance of the paretic UL (9).

The analysis of cultural invariance is considered the formal way to examine the cross-cultural validity of a scale (16). This analysis has been recommended (15, 16), because it is possible to have versions of scales that are valid in a given country, but work differently in another, compromising their cross-cultural use (16). In this study, both the invariance of estimates of item difficulty (DIF analysis) and estimates of person's ability were examined. The results of the DIF analysis indicated differences between the ABILHAND-Original and ABILHAND-Brazil calibrations. Other studies, which also analysed the DIF between countries, attributed the differences in the estimates of item difficulty to cultural aspects, such as traditions, lifestyle habits, and environmental contexts, such as eating situation and type of food (15, 17, 18). However, in the present study, these differences in the item calibrations did not impact the measure, because, regardless of the calibration, the manual ability estimates were the same, supporting the cultural invariance of the ABILHAND. This finding is particularly important because, given the context-dependent nature of the estimates of human abilities, the invariance is the exception, not the rule (14).

These findings suggest that the cross-cultural validity of the ABILHAND further supports its construct validity, since a scale should work in the same way to measure the same construct within different contexts (14, 15). These results have some practical implications. They suggest that the ABILHAND-Original and ABILHAND-Brazil calibrations could be used interchangeably. Thus, Rasch analysis (available from www.rehab-scales.org) based on the ABILHAND-Original calibration, can be used for the conversion of the ordinal scores of the ABILHAND-Brazil into linear measures of manual ability. This, in addition to facilitating the use of the ABILHAND within both clinical and research contexts in Brazil, will allow its use in international/multicentric studies.

The strength of the present study was that the item calibration was stable, even though the sample demonstrated large variability regarding the time since onset of stroke, levels of motor recovery, and previous dominance of the paretic UL. From a measurement perspective, this provides evidence of the external validity of the ABILHAND. However, within a clinical perspective, the inclusion of more homogenous samples would be useful to tailor interventions to specific needs. Penta et al. (9), by means of DIF analyses, showed that the item calibration of the ABILHAND was stable across groups with various clinical characteristics, such as the paretic side (dominant vs non-dominant) and time since stroke onset (<2 vs ≥ 2 years). However, their sample was relatively small for strong conclusions regarding sub-group analyses. Future studies should examine these issues in larger samples, in order to obtain more powerful results.

It is important to point out that, from a list of 485 contacts, only 121 agreed to participate and 107 met the inclusion criteria. It is well known that sample size depends on financial support, time, and availability of volunteers. Despite difficulties in recruitment, the targeting of the sample was good and, thus, the current sample was sufficient to give a degree of precision for the estimates of ± 0.5 logits (39).

This study should be considered as a preliminary analysis of the cross-cultural validity of the ABILHAND, since the generalizability of these results is limited to Brazil, Belgium, and Italy. Future studies should include samples from other countries for the cumulative evidence of the cross-cultural validity of the ABILHAND.

In conclusion, the results of this study suggest that the ABILHAND for stroke individuals demonstrated cross-cultural validity and satisfactory measurement properties for use within both clinical and research contexts in Brazil.

ACKNOWLEDGEMENTS

Financial support was provided by the Brazilian funding agencies: CAPES, CNPq, and FAPEMIG.

REFERENCES

1. Martins SC, Pontes-Neto OM, Alves CV, de Freitas GR, Filho JO, Tosta ED, et al. Past, present, and future of stroke in middle-income countries: the Brazilian experience. Int J Stroke 2013; 8 Suppl A100: 106–111.
2. Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA, et al. Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. Lancet 2014; 383: 245–255.
3. Minelli C, Fen LF, Minelli DP. Stroke incidence, prognosis, 30-day, and 1-year case fatality rates in Matão, Brazil: a population-based prospective study. Stroke 2007; 38: 2906–2911.
4. Harris JE, Eng JJ. Paretic upper-limb strength best explains arm activity in people with stroke. Phys Ther 2007; 87: 88–97.
5. Faria-Fortini I, Michaelsen SM, Cassiano JG, Teixeira-Salmela LF. Upper extremity function in stroke subjects: relationships between the International Classification of Functioning, Disability, and Health domains. J Hand Ther 2011; 24: 257–264.
6. Persson HC, Parziali M, Danielsson A, Sunnerhagen KS. Outcome and upper extremity function within 72 hours after first occasion of stroke in an unselected population at a stroke unit. A part of the SALGOT study. BMC Neurol [serial on the Internet]. 2012 Dec [cited 2015 Jan 7]; 12(162): [about 6p.]. Available from: http://www.biomedcentral.com/1471-2377/12/162.
7. Hunter SM, Crome P. Hand function and stroke. Rev Clin Gerontol 2002; 12: 68–81.
8. World Health Organization. International Classification of Functioning, Disability and Health. Geneva: World Health Organization; 2001.
9. 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: 1627–1634.
10. Batcho CS, Tennant A, Thonnard JL. ACTIVLIM-Stroke: a cross-cultural Rasch-built scale of activity limitations in patients with stroke. Stroke 2012; 43: 815–823.
11. Ekstrand E, Lindgren I, Lexell J, Brogårdh C. Test-retest reliability of the ABILHAND Questionnaire in persons with chronic stroke. PM R 2014; 6: 324–331.
12. Connell LA, Tyson SF. Clinical reality of measuring upper-limb ability in neurologic conditions: a systematic review. Arch Phys Med Rehabil 2012; 93: 221–228.
13. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: time to end malpractice? J Rehabil Med 2012; 44: 97–98.
14. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. 2nd ed. New York: Routledge; 2010.
15. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004; 42 Suppl 1: I37–I48.
16. Küçükdeveci AA, Sahin H, Ataman S, Griffiths B, Tennant A. Issues in cross-cultural validity: Example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. Arthritis Rheum 2004; 51: 14–19.
17. Chien CW, Brown T, McDonald R. Cross-cultural validity of a naturalistic observational assessment of children´s hand skills: a study using Rasch analysis. J Rehabil Med 2011; 43: 631–637.
18. Lundgren-Nilsson A, Grimby G, Ring H, Tesio L, Lawton G, Slade A, et al. Cross-cultural validity of the Functional Independence

Measure items in stroke: a study using Rasch analysis. J Rehabil Med 2005; 37: 23–31.
19. Chien CW, Bond TG. Measurement properties of fine motor scale of Peabody developmental motor scales-second edition: a Rasch analysis. Am J Phys Med Rehab 2009; 88: 376–386.
20. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007; 57: 1358–1362.
21. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self report measures. Spine 2000; 25: 3186–3191.
22. Bertolucci PHF, Brucki SMD, Campacci SR, Juliano Y. The Mini-Mental State Examination in an out-patient population: influence of literacy. Arq Neuro Psiquiatr 1994; 52: 1–7.
23. Gregson JM, Leathley M, Moore AP, Sharma AK, Smith TL, Watkins CL. Reliability of the Tone Assessment Scale and the modified Ashworth Scale as clinical tools for assessing poststroke spasticity. Arch Phys Med Rehabil 1999; 80: 1013–1016.
24. Bohannon RW. Test-retest reliability of hand-held dynamometry during a single session of strength assessment. Phys Ther 1986; 66: 206–209.
25. Faria CDCM, Aguiar LT, Lara EM, Souza LAC, Martins JC, Teixeira-Salmela LF. Dynamometry for the assessment of grip, pinch, and trunk strength in subjects with chronic stroke: reliability and various sources of outcome values. Int J Phys Med Rehabil [serial on the Internet]. 2013 Nov [cited 2015 Jan 7]; 1 (168): [about 5p]. Available from: http://omicsonline.org/dynamometry-for-the-assessment-of-grip-pinch-and-trunk-strength-in-subjects-with-chronic-stroke-2329-9096.1000168.php?aid=21367.
26. Université catholique de Louvain. Rehab-scales.org [Internet]. 2007 [cited 2015 Jan 7]. Available from: http://www.rehab-scales.org/evaluation-scales-in-rehabilitation.html.
27. Wild D, Grove A, Martin M, Erenmenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health 2005; 8: 94–104.
28. Maki T, Quagliato EMAB, Cacho EWA, Paz LPS, Nascimento NH, Inoue MMEA, et al. Reliability study on the application of the Fugl-Meyer scale in Brazil. Braz J Phys Ther 2006; 10: 177–183.
29. Michaelsen SM, Rocha AS, Knabben RJ, Rodrigues LP, Fernandes CG. Translation, adaptation and inter-rater reliability of the administration manual for the Fugl-Meyer assessment. Braz J Phys Ther 2011; 15: 80–88.
30. Linacre JM. A User's guide to Winsteps® Ministep Rasch-model computer programs. Program manual 3.81.0. Chicago: Winsteps.com; 2014 [cited 2015 Apr 9]. Available from: http://www.winsteps.com/winman/index.htm.
31. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003; 35: 105–115.
32. Riazi A, Aspden T, Jones F. Stroke self-efficacy questionnaire: a Rasch-refined measure of confidence post stroke. J Rehabil Med 2014; 46: 406–412.
33. Smith RM, Schumacker RE, Bush MJ. Using item mean squares to evaluate fit to the Rasch model. J Outcome Meas 1998; 2: 66–78.
34. Chern JS, Kielhofner G, de las Heras CG, Magalhães LC. The volitional questionnaire: psychometric development and practical use. Am J Occup Ther 1996; 50: 516–525.
35. Raîche G. Critical eigenvalue sizes (variances) in standardized residual principal components analysis. Rasch Meas Trans 2005; 19: 1012.
36. Pomeroy IM, Tennant A, Young CA. Rasch analysis of the WHOQOL-BREF in post polio syndrome. J Rehabil Med 2013; 45: 873–880.
37. Portney LG, Watkins MP. Foundations of clinical research: application to practice. 3rd ed. Upper Saddle River: Prentice-Hall; 2008.
38. Miller KJ, Slade AL, Pallant JF, Galea MP. Evaluation of the psychometric properties of the upper limb subscales of the Motor Assessment Scale using a Rasch analysis model. J Rehabil Med 2010; 42: 315–322.
39. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans 1994; 7: 328.