

Assessment of the measurement properties of the post-stroke motor function instruments available in Brazil: a systematic review

Elaine Lima¹, Luci F. Teixeira-Salmela², Luan Simões¹,
Ana C. C. Guerra¹, Andrea Lemos¹

ABSTRACT | Background: While there are several instruments in Brazil that measure motor function in patients after stroke, it is unknown whether the measurement properties of these instruments are appropriate. **Objective:** To identify the motor function instruments available in Brazil for patients after stroke. To assess the methodological quality of the studies and the results related to the measurement properties of these instruments. **Method:** Two independent reviewers conducted searches on PubMed, LILACS, CINAHL, Web of Science, and Scopus. Studies that aimed to cross-culturally adapt an existing instrument or create a Brazilian instrument and test at least one measurement property related to motor function in patients after stroke were included. The methodological quality of these studies was checked by the COSMIN checklist with 4-point rating scale and the results of the measurement properties were analyzed by the criteria developed by Terwee et al. **Results:** A total of 11 instruments were considered eligible, none of which were created in Brazil. The process of cross-cultural adaptation was inadequate in 10 out of 11 instruments due to the lack of back-translation or due to inappropriate target population. All of the instruments presented flaws in the measurement properties, especially reliability, internal consistency, and construct validity. **Conclusion:** The flaws observed in both cross-cultural adaptation process and testing measurement properties make the results inconclusive on the validity of the available instruments. Adequate procedures of cross-cultural adaptation and measurement properties of these instruments are strongly needed.

Keywords: stroke; validity of tests; reproducibility of results; translating; physical therapy.

BULLET POINTS

- 11 studies were found that assessed post-stroke motor function in Brazilian patients.
- Most of the cross-cultural adaptation was conducted without the target population.
- Flaws in the measurement properties made the results inconclusive.
- Caution should be taken in the selection of instruments for research and clinical practice.

HOW TO CITE THIS ARTICLE

Lima E, Teixeira-Salmela LF, Simões L, Guerra ACC, Lemos A. Assessment of the measurement properties of the post-stroke motor function instruments available in Brazil: a systematic review. *Braz J Phys Ther.* 2016 Mar-Apr; 20(2):114-125. <http://dx.doi.org/10.1590/bjpt-rbf.2014.0144>

● Introduction

Various measurement instruments have been created with the objective of assessing motor function in post-stroke individuals¹⁻³. These instruments aim to verify the ability to maintain or change the body's position in space, walk and move around, move and handle objects, as well as verify motor coordination and fine manual motricity¹⁻³. These abilities involve aspects related to activities and participation and the structure and function of the organs and systems, as described in the International Classification of Functioning, Disability, and Health (ICF)⁴.

The application of these instruments aims to measure upper limb function, trunk function, or global motor function¹⁻³. Some instruments assess performance through the observation of performed activities, while others are based on questionnaires on motor function¹⁻³. After stroke, motor function can present various degrees of impairment and generate social and economic loss. Therefore, it is essential to use valid instruments to achieve an effective rehabilitation⁵⁻⁷.

In general, the instruments used in Brazil to assess post-stroke motor function were developed in other

¹Departamento de Fisioterapia, Universidade Federal de Pernambuco (UFPE), Recife, PE, Brazil

²Departamento de Educação Física, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil

Received: Apr. 16, 2015 Revised: Aug. 12, 2015 Accepted: Aug. 26, 2015

countries, usually in English and, consequently, targeted to the original population^{1,8}. However, before an instrument can be used in a new country, culture, and/or language, a cross-cultural adaptation process is necessary. This process requires a standardized method involving the language translation and the cross-cultural adaptation to maintain its content validity^{9,10}. After this process, the new scale should be applied to the new target population and its measurement properties can be analyzed to check if the adapted instrument truly measures the construct in the new setting⁹⁻¹².

The instrument can only be considered valid and reliable for use in a new cultural-clinical context through the adequate evaluation of the measurement properties⁹⁻¹¹. The objectives of this systematic review were to identify the measurement instruments of motor functions in post-stroke individuals available in Brazil,

to assess the methodological quality of the studies, and to assess the results of these studies.

● Method

Two independent reviewers (EL and LS) conducted searches and selected eligible studies in the PUBMED, LILACS, SCOPUS, CINAHL, and WEB OF SCIENCE databases between February and March of 2014, according to the search strategy presented in Table 1. There was no language restriction.

Either cross-cultural adaptation studies or Brazilian instruments that assessed the motor function of post-stroke individuals in at least one item were considered eligible. Furthermore, these studies had to have verified at least one measurement property of

Table 1. Research strategies for each research database.

DATABASE	RESEARCH STRATEGY
MEDLINE (PUBMED)	((“Brazil “ [Mesh]) OR Portuguese OR Brazilian) AND ((“Stroke” [Mesh]) OR (“Paresis” [Mesh])) AND ((“Questionnaires” [Mesh]) OR scale OR test OR performance based test) AND Sensitive search filter for measurement properties NOT Exclusion Filter *
LILACS	(Brazil OR Portuguese OR Brazilian OR) AND (Stroke OR Stroke OR paresis) AND (Trunk OR upper limbs OR lower limbs OR sensorimotor function OR motor OR Function motor activity OR mobility OR coordination balance OR instrumentation OR comparative studies OR validation studies OR translations OR translation adjustment OR cross-cultural equivalence OR Validity OR validation OR Reliability OR reproducibility OR reproducible OR psychometric tests OR psychometric properties OR clinimetric clinimetric OR property OR valuation OR inter-observer OR variation results OR Intra-examiner OR many retest OR inter-rater OR intraobserver OR interparticipants OR intraparticipants OR internal consistency Rasch OR Effect OR Effect floor ceiling OR disability assessment OR questionnaires OR scale tests)
CINAHL AND WEB OF SCIENCE	(Brazil OR Brazilian OR Portuguese) AND (Stroke OR Paresis) AND (questionnaires OR scale OR test OR comparative studies OR validation studies OR validation OR translations OR cross cultural OR cross cultural adaptation OR cross cultural comparison OR cross-cultural equivalence OR validity OR reliability OR reproducibility OR psychometrics OR clinimetrics OR outcome assessment OR observer variation OR reproducibility of results OR internal consistency OR alpha Cronbach OR agreement OR precision OR test-retest OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intraobserver OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR concordance OR intraclass OR dimension OR subscale OR responsivity OR ceiling effect OR floor effect OR Item response model OR IRT OR Rasch)
SCOPUS	TITLE-ABS-KEY(stroke OR paresis) AND TITLE-ABS-KEY(cross cultural adaptation OR cross cultural comparison OR translation) AND TITLE-ABS-KEY(Brazil OR Brazilian OR Portuguese) TITLE-ABS-KEY(stroke OR paresis) AND TITLE-ABS-KEY(validation OR validity OR validation studies) AND TITLE-ABS-KEY(Brazil OR Brazilian OR Portuguese) TITLE-ABS-KEY(stroke OR paresis) AND TITLE-ABS-KEY(psychometrics OR clinimetrics OR reliability OR reproducibility OR responsiveness OR internal consistency OR intra examiner OR inter examiner OR responsivity OR ceiling effect OR floor effect) AND TITLE-ABS-KEY(brazil OR brazilian OR portuguese)

*In accordance with the recommendations of the Consensus-based Standards for the selection of health Measurements Instruments-COSMIN¹³.

these instruments. Studies that involved individuals with other neurological conditions were excluded.

The two reviewers (EL and LS) screened the studies by title and abstract performing a pre-selection through eligibility criteria on the computer screen. Then, they read the full text of the studies potentially eligible to confirm their inclusion. It was pre-defined that disagreements between two reviewers were arbitrated by a third reviewer (AL).

The data extraction was performed in a standardized way through a pre-established data extraction form. The following data were extracted: title, authors, year of publication, journal, study objectives, eligibility criteria of the participants, instrument objective (discriminative, predictive or evaluative)¹², number of subscales/items/domains, and domain assessed according to ICF⁴.

The evaluation of the methodological quality of the included studies was performed through the COSMIN checklist with 4-point rating scale, which is a tool created through the Consensus-based Standards for the selection of health Measurements Instruments (COSMIN), with the aim of scoring and classifying the quality of the methods used for the study of each measurement property¹⁴⁻¹⁶.

The COSMIN checklist with 4-point rating scale is composed of nine boxes: A- Internal consistency, B- Reliability, C- Measurement error, D- Content validity, E- Structural validity, F- Hypothesis tests, G- Cross-cultural validity, H- Criterion validity, and I- Responsiveness¹⁴⁻¹⁶.

Each box includes a series of items that assess the measurement property methodology. These items are classified on a scale of 4 points: 1- Poor, 2- Fair, 3- Good, and 4- Excellent. The final classification for each box is determined by the lowest score achieved by any of the items¹⁴⁻¹⁶.

In addition to the boxes mentioned above, there is still another box that should be completed for each measurement property. This box aims to identify the clinical-epidemiological profile of the population, analyzing age mean, distribution by gender, illness characteristic, country of origin, and spoken language¹⁴⁻¹⁶.

For example, to assess the internal consistency, box A presents 11 items: the first 3 items assess the missing data. Item 4 assesses the sample size; items 5, 6, and 7 assess questions related to unidimensionality; item 8 verifies the presence of other methodological flaws; and the other items verify the statistical method¹⁴⁻¹⁶.

Item 4 of this box assesses the sample size for internal consistency as follows: Excellent (N=100),

Good (between 50 and 99 participants), Fair (between 30 and 40 participants), and Poor (less than 30 participants). Similar to item 4, the remaining items are scored on a 4-point scale, according to specific criteria. In the end, even if the instrument has obtained “excellent” classification in the other items, but in item 4 received a “good” score for having a sample size between 50 and 99 participants, the internal consistency of the instrument will be classified as having “good” internal consistency as the lowest score is used¹⁴⁻¹⁶.

Furthermore, the COSMIN recommends that to complement the evaluation of an instrument, the quality criteria developed by Terwee et al.¹¹ should be used; these criteria classify the measurement properties as Positive (+), Negative (-), or doubtful (?) focusing on the analysis of the obtained results¹¹. The use of the Terwee et al.¹¹ criteria complements the evaluation of the measurement properties, as the COSMIN does not determine the cut offs that are considered adequate for the statistical analysis of each measurement property. In other words, the fact that a study used Cronbach’s α , one of the statistical measurements advocated by COSMIN, to verify the internal consistency does not guarantee the quality of this property, as adequate values may not have been reached^{11,14-16}.

For example, internal consistency receives a positive score when the unidimensionality is verified, with the participation of 100 or more individuals and through Cronbach’s α (between 0.70 and 0.95). If α does not reach this interval, the score will be negative. When the unidimensionality is not verified, or if there is another methodological flaw, the score will be classified as inconclusive¹¹.

● Results

A total of 529 studies were found, of which only 14 studies¹⁷⁻³⁰ were included through the eligibility criteria (Figure 1). Two instruments (Test Évaluant les Membres supérieurs des Personnes Âgées - TEMPA²⁰ and the Jebsen-Taylor Test²¹) were not specifically created for post-stroke individuals; however, they have been validated in Brazil for this population and were included in this review.

In the 14 studies included, 11 instruments were identified. Three of them (Motor Activity Log - MAL^{18,19}, Fugl-Meyer scale^{23,24}, and National Institute of Health Stroke Scale - NIHSS^{25,26}) were analyzed in two studies each, and the other 8 in only one

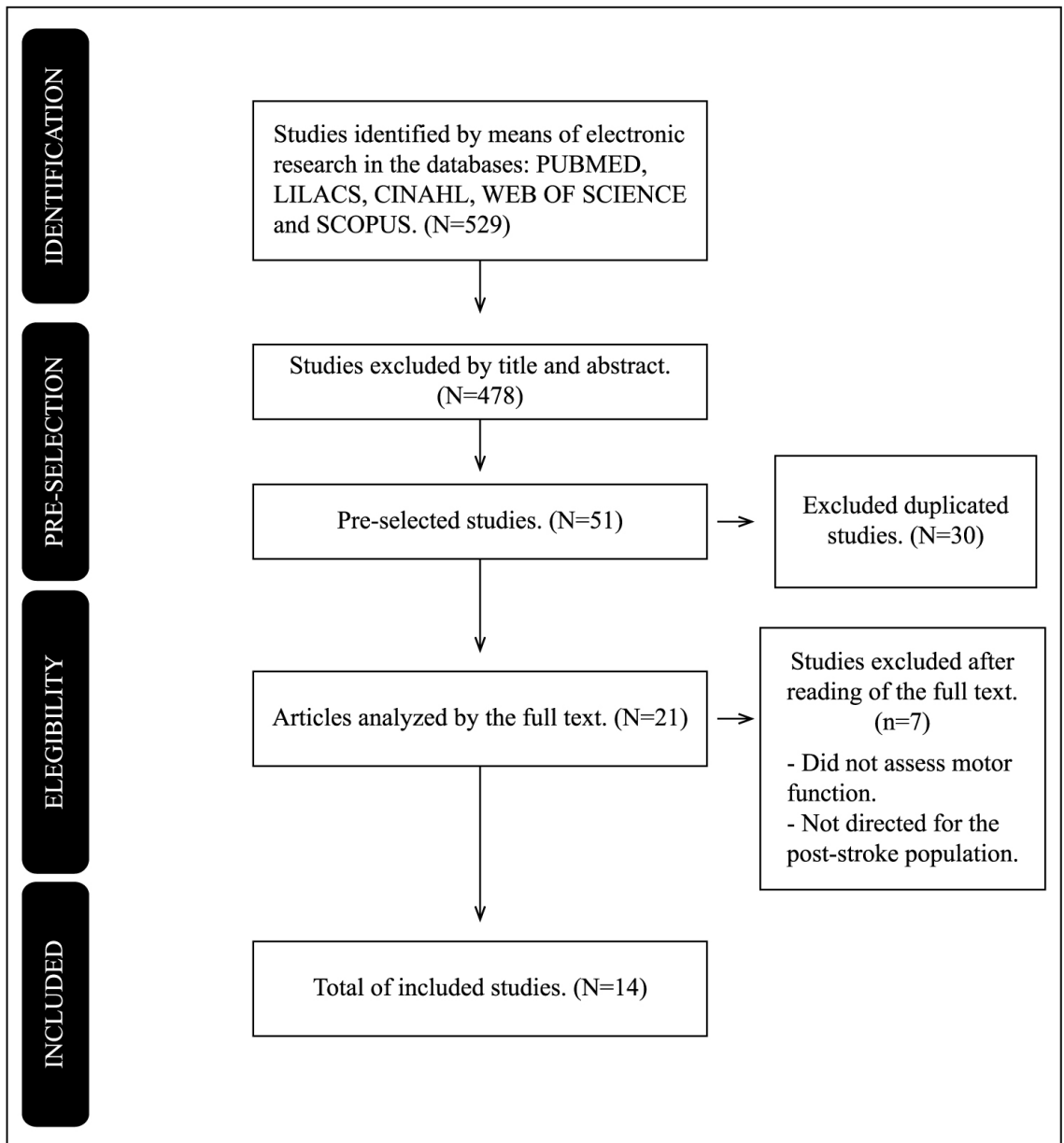


Figure 1. Identification, selection, and inclusion of the studies.

study. The characterization of the included studies is presented in Table 2.

None of the included instruments were Brazilian, therefore all had to be submitted to the cross-cultural adaptation process. The most frequently measured properties were reliability ($n=11$ studies), construct validity through hypothesis testing ($n=6$ studies), and internal consistency ($n=6$ studies). None of the included studies assessed responsiveness. The evaluation of the measurement properties is shown in Table 3.

Cross-cultural adaptation

Only one instrument (MAL)^{18,19} was adapted to Brazilian culture in accordance with the recommended method^{9,14-16}. The cross-cultural adaptation of the Rivermead Mobility Index (RMI)²² obtained a “good” classification, given that it is not clear whether an expert committee participated or whether a pre-test was conducted⁹.

Six instruments^{17,21,23,25-27,30} obtained a classification considered “poor” for the cross-cultural adaptation

Table 2. Description of the instruments that assess the motor function of post-stroke individuals available in Brazil.

INSTRUMENT	AUTHOR/YEAR	CONSTRUCT	OBJECTIVE	TYPE OF INSTRUMENT	DIMENSION OF CIF	N° OF ITEMS/SUBSCALES	BRAZIL VERSION (AUTHOR/YEAR)
Postural Assessment Scale	Benaim et al. ³¹ (1999)	Maintenance and postural changes	Evaluative	Based on performance	Activity	12 items	Yoneiama et al. ²⁸ (2008)
Trunk impairment Scale	Fujiwara et al. ³² (2004)	Trunk motor function	Evaluative	Based on performance	Structure and function	7 items	Lima et al. ³⁰ (2008)
Trunk impairment Scale	Verheyden et al. ³³ (2004)	Trunk motor function	Evaluative	Based on performance	Structure and function /Activity	17 items	Castelassi et al. ²⁹ (2009)
Fugl-Meyer Scale	Fugl-Meyer et al. ³⁴ (1975)	Global sensory-motor function	Evaluative	Based on performance	Structure and function /Activity	7 subscales	Maki et al. ²⁴ (2006) Michaelsen et al. ²³ (2011)
Jebesen-Taylor Test	Jebesen et al. ³⁵ (1969)	Upper limb motor function	Evaluative	Based on performance	Activity	7 items	Ferreiro et al. ²¹ (2010)
Motor Assessment Scale (MAS)	Carr et al. ³⁶ (1985)	Global motor function	Evaluative	Based on performance	Activity	8 items	Conte et al. ²⁷ (2009)
Motor Activity Log (MAL)	Uswatte et al. ³⁷ (2006)	Upper limb motor function	Evaluative	Questionnaire	Participation	30 items	Saliba et al. ¹⁸ (2011) Pereira et al. ¹⁹ (2012)
National Institutes of Health stroke scale (NIHSS)	Brott et al. ³⁸ (1989)	Losses from the Stroke	Predictive	Based on performance	Activity	11 items (2 for motor function)	Cincura et al. ²⁵ (2009) Caneda et al. ²⁶ (2006)
Rivermead Mobility Index (RMI)	Collen et al. ³⁹ (1991)	Global motor function	Evaluative	Questionnaire	Activity	15 items	Pavan et al. ²² (2010)
TEMPA (Test Évaluant les Membres supérieurs des Personnes Âgées)	Desrosiers et al. ⁴⁰ (1993)	Upper limb motor function	Evaluative	Based on performance	Activity	9 items (8 items in the Brazilian version)	Michaelsen et al. ²⁰ (2008)
Wolf Motor Function Test	Morris et al. ⁴¹ (2004)	Upper limb motor function	Evaluative	Based on performance	Structure and function /Activity	17 items	Pereira et al. ¹⁷ (2011)

Table 3. Evaluation of the methodological quality of the included studies through the COSMIN checklist with 4-point rating scale: Consensus-based standard for the selection of health measurement instruments.

INSTRUMENT	AUTHOR/YEAR	USED TRI	INTERNAL CONSISTENCY (A)	RELIABILITY (B)	MEASUREMENT ERROR (C)	F HYPOTHESIS TEST (F)	CROSS-CULTURAL ADAPTATION (G)	GENERALIZATION BY BOX
Postural Assessment Scale	Yonejama et al. ²⁸ (2008)	No	Poor	(Inter and test-retest) Poor	-	Poor	Fair	A Excellent B Excellent C Excellent F Excellent G Poor
Trunk Commitment Scale	Lima et al. ³⁰ (2008)	No	Poor	(Inter and test-retest) Poor	-	Poor	Poor	A Excellent B Excellent F Excellent G Poor
Trunk Deficiency Scale	Castelassi et al. ²⁹ (2009)	No	Poor	Poor	Poor	Poor	Fair	A Excellent B Excellent F Excellent G Reasonable
Fugl-Meyer Scale	Maki et al. ²⁴ (2006)	No	-	(Inter and test-retest) Poor	-	-	Fair	B Excellent G Poor
Fugl-Meyer Scale (manual)	Michaelsen et al. ²³ (2011)	No	-	(Inter) Poor	-	-	Poor	B Excellent G Poor
Jebesen-Taylor Test	Ferreiro et al. ²¹ (2010)	No	Poor	(Inter and Intra) Fair	-	-	Poor	A Excellent B Excellent G Poor
Motor Assessment Scale	Conte et al. ²⁷ (2009)	No	-	(Inter and Intra) Poor	-	-	Poor	B Excellent G Poor
Motor Activity Log	Saliba et al. ¹⁸ (2011)	Yes	Good	Poor (test-retest)	-	-	Excellent	A Excellent B Excellent F Excellent G Excellent
Motor Activity Log	Pereira et al. ¹⁹ (2012)	No	-	(Inter and test-retest) Poor	(Inter and test-retest) Poor	Fair	-	B Excellent C Excellent F Excellent

TRI: Theory of the Answer to the TEMPA item; Test Évaluant les Membres supérieurs des Personnes Âgées. The following properties were not analyzed in the studies: D: Content Validity; E: Structural validity; H: Criteria validity; I: Responsiveness; J: Interpretability.

Table 3. Continued...

INSTRUMENT	AUTHOR/YEAR	USED TRI	INTERNAL CONSISTENCY (A)	RELIABILITY (B)	MEASUREMENT ERROR (C)	F HYPOTHESIS TEST (F)	CROSS-CULTURAL ADAPTATION (G)	GENERALIZATION BY BOX
National Institute of Health stroke scale	Cincura et al. ²⁵ (2009)	No	-	Good	-	Fair	Poor	B Excellent F Excellent G Poor
National Institute of Health stroke scale	Caneda et al. ²⁶ (2006)	No	-	Good	-	-	Poor	B Excellent
Rivermead Mobility Index	Pavan et al. ²² (2010)	No	Poor	Poor (test-retest)	-	-	Good	A Excellent B Excellent G Poor
TEMPA	Michaelsen et al. ²⁰ (2008)	No	-	(Inter and test-retest) Poor	-	Poor	Poor	B Excellent F Excellent G Poor
Wolf Motor Function Test	Pereira et al. ¹⁷ (2011)	No	-	(Inter and Intra) Poor	(Inter and Intra) Poor	-	Poor	B Excellent C Excellent G Poor

TRI: Theory of the Answer to the TEMPA item; Test Évaluant les Membres supérieurs des Personnes Âgées. The following properties were not analyzed in the studies: D: Content Validity; E: Structural validity; H: Criteria validity; I: Responsiveness; J: Interpretability.

process. The TEMPA²⁰ and Motor Assessment Scale (MAS)²⁷ were created through simple translations into Portuguese. The NIHSS^{25,26} was inadequately adapted in two studies, one of which only performed a single translation into Portuguese²⁵. In the adaptation of the Wolf Motor Test (Wolf)¹⁷, Jebsen-Taylor Test²¹, and the Trunk Impairment Scale³⁰, pre-tests were not performed either.

For the Fugl-Meyer Scale^{23,24}, Posture Assessment Scale²⁸, and Trunk Deficiency Scale²⁹, the quality of the adaptation process was “fair” because it included a pre-test but did not include an adequate description of the assessed sample. However, the Fugl-Meyer Scale²³ manual, which was produced in a different study to the production of the instrument, presented a “poor” process as it included only one translation into Portuguese. In the Posture Assessment Scale²⁸ and Trunk impairment Scale²⁹, the translation and back translation were performed by only one translator.

Reliability

All of the instruments were tested for reliability. Eight (MAS²⁷, MAL^{18,19}, Wolf Motor Function Test¹⁷, TEMPA²⁰, Posture Assessment Scale²⁸, Trunk impairment Scale³⁰, and the study of the Fugl-Meyer manual²⁴) received a “poor” classification because they included fewer than 30 participants and used the intraclass correlation coefficient (ICC) when this was not indicated. The studies of the Fugl-Meyer Scale²³ (N=50) and RMI²² (N=95) had good samples but were classified as “poor” for having used inadequate statistical methods (i.e. ICC and the Wilcoxon test, respectively). The Jebsen-Taylor Test²¹ was considered “fair” for presenting a sample between 30 and 49 individuals (n=40). The reliability of the NIHSS was verified in two studies with “good” methodology and samples of 51 and 62 participants, respectively^{25,26}.

Measurement error

The measurement error was verified in three instruments (Trunk impairment Scale²⁹, Wolf Motor Function Test¹⁷, and MAL^{18,19}) through the Bland-Altman plot analysis; however, the methodological quality was classified as “poor” because the sample included less than 30 individuals.

Internal consistency

Six instruments^{18,19,21,22,28-30} were tested for internal consistency; however, the methodological quality was classified as “poor” in all of them. In five instruments

(Posture Assessment Scale²⁸, RMI²², Jebsen Taylor Test²¹, and Trunk impairment Scales^{29,30}), the reason was the lack of factor analysis. Moreover, in the Posture Assessment Scale²⁸ and Trunk impairment Scales^{29,30}, the sample included less than 30 individuals and in the study of MAL^{18,19}, the sample included less than 5 individuals per item of the instrument for unidimensionality.

Construct validity

Construct validity was analyzed in six instruments (MAL¹⁹, TEMPA²⁰, Posture Assessment Scale²⁸, NIHSS²⁵, and Trunk Impairment Scales^{29,30}) through the hypothesis tests by correlation with the Fugl-Meyer Scale^{19,20,28,29}, Barthel Index²⁵, Berg Balance Scale, and Functional Independence Measure³⁰. The study method used in four of these instruments was classified as “poor” due to inadequate sample size (n<30)^{20,28-30}.

The MAL¹⁹ and NIHSS^{25,26} presented “fair” methodological quality in the validity tests, as the hypotheses about the direction and magnitude of the correlation were not previously formulated or described in the study; however, it was possible to assume the expected direction for the correlation (positive or negative).

Terwee criteria

As for the evaluation of the results of the measurement property analysis using the criteria of Terwee et al.¹¹, the majority of the studies presented doubtful results in the study of measurement properties, with the exception of the inter-examiner reliability of the NIHSS^{25,26}, which presented positive results with Kappa coefficient >0.70 in items 5a, 5b, 6a, and 6b (referring to upper and lower limb motor function) (Table 4).

The results of the measurement error tests of the internal consistency and of construct validity were considered doubtful due to the methodological flaws presented, as described previously^{17,19,28}.

The ceiling and floor effects, which reflect interpretability, were verified in two instruments (Trunk Impairment Scale²⁹ and Posture Assessment Scale²⁸). The percentage of individuals who reached the minimum and maximum scores was lower than 15%, but with an inadequate sample size (<50). However, other measures of interpretability like the minimum clinically important difference and minimum important difference were not analyzed. Finally, criterion validity and responsiveness were not tested in any of the eligible studies.

Table 4. Measurement properties assessment through the Terwee et al.¹¹ criteria (2007).

Instrument	Author/ year	Internal Consistency	Reliability	Measurement error	Content validity	Construct validity	Criteria validity	Responsiveness	Floor and ceiling effect	Interpretability
Posture Assessment Scale	Yoneima et al. ²⁸ (2008)	?	(inter and test-retest) ?	NA	NA	?	NA	NA	?	NA
Trunk commitment Scale	Lima et al. ³⁰ (2008)	?	?	NA	NA	?	NA	NA	NA	NA
Trunk Deficiency Scale	Castelassi et al. ²⁹ (2009)	?	?	?	NA	?	NA	NA	?	NA
Fugl-Meyer Scale	Maki et al. ²⁴ (2006)	NA	(Inter and test-retest) ?	NA	NA	NA	NA	NA	NA	NA
Fugl-Meyer Scale Manual	Michaelsen et al. ²³ (2011)	NA	(Inter) ?	NA	NA	NA	NA	NA	NA	NA
Teste de Jebsen-Taylor	Ferreiro et al. ²¹ (2010)	?	(inter and intra) ?	NA	NA	NA	NA	NA	NA	NA
Motor Assessment Scale	Conte et al. ²⁷ (2009)	NA	(inter and intra) ?	NA	NA	NA	NA	NA	NA	NA
Motor Activity Log	Saliba et al. ¹⁸ (2011)	?	(test-retest) ?	NA	NA	?	NA	NA	NA	NA
Motor Activity Log	Pereira et al. ¹⁹ (2012)	NA	(inter and test-retest) ?	(inter and test-retest) ?	NA	?	NA	NA	NA	NA
NIHSS	Cincura et al. ²⁵ (2009)	NA	(inter) +	NA	NA	?	NA	NA	NA	NA
NIHSS	Caneda et al. ²⁶ (2006)	NA	(inter) +	NA	NA	NA	NA	NA	NA	NA
Rivermead Mobility Index	Pavan et al. ²² (2010)	?	(test-retest) ?	NA	NA	NA	NA	NA	NA	NA
TEMPA	Michaelsen et al. ²⁰ (2008)	NA	(inter and test-retest) ?	NA	NA	?	NA	NA	NA	NA
Wolf Motor Function Test	Pereira et al. ¹⁷ (2011)	NA	(Inter and intra) ?	(Inter and intra) ?	NA	NA	NA	NA	NA	NA

(+)= Positive; (-)= Negative; (?)= Inconclusive; (NA)= Not Assessed. TEMPA: Test Évaluant les Membres supérieurs des Personnes Âgées; NIHSS: National Institute of Health Stroke Scale.

● Discussion

The results of this review showed that the available instruments in Brazil for assessing post-stroke motor function are arising from cross-cultural adaptation, not from newly developed Brazilian. However, the findings are inconclusive regarding the quality of the cross-cultural adaptation as well as from measurement properties, due to flaws with regards to methodology. The main methodological flaw observed during the cross-cultural adaptation process of the included instruments was the absence of a pre-testing of the final version^{17,20-22,27,30}. Only one instrument (Motor Activity Log - MAL)^{18,19} followed the recommended processes for an adequate cross-cultural adaptation.

The goal of applying the instrument in the target population (pre-test) before the measurement property analysis aims to identify possible imperfections in the interpretation of the items of an interview and the viability of the tasks proposed by the instrument for the target population. Therefore, the performance of the pre-test allows the identification of possible adjustments necessary in the instrument, based on the direct participation of the population for which it was adapted¹¹.

Although some instruments performed a pre-test, most of the studies did not described the sample properly^{24,28,29}. To allow the generalization of the results of a cross-cultural adaptation, the COSMIN checklist recommends that the participants involved in the pre-test should be clinically and epidemiologically reported in terms of age, gender, characteristics of the illness, and source of patients (hospital, clinic, community, etc.)¹⁴⁻¹⁶.

The absence of a back translation was also verified in some instruments^{20,27}. This stage has the important aspect of allowing the verification of semantic equivalence between the original instrument and what was created in the new language, allowing necessary adjustments in the new version. It was also observed that, in some instruments^{28,29}, the stages of translation and back translation were performed by a single translator. The performance of multiple translations is recommended in the literature because it allows the interaction between specialists in the construct and in the languages involved, allowing a more adequate process of cultural adaptation and the maintenance of semantic equivalence¹⁴⁻¹⁶.

Concerning the measurement properties, methodological flaws were also verified. The reliability was verified in all studies; however, in the majority of these, a sample size of less than 30 participants was

selected^{17-20,22,27-30}. The adequate number is at least 50 participants, and for an ideal sample, the recruitment of at least 100 participants is recommended¹⁴⁻¹⁶.

In addition, the intra-lass correlation coefficient was often chosen as the statistical method when it was, in fact, inadequate. The adequate method for instruments with ordinal type scores is the Kappa coefficient^{17-20,22-24,27-30}. The only instrument with an adequate study method for reliability, the NHSS, presented flaws in the cross-cultural adaptation^{25,26}.

For internal consistency, the majority of the studies did not report factorial analysis or unidimensionality study of the items^{21,22,28-30}. These analyses are important because they intend to verify the number of dimensions into which the items are distributed and whether subscales are formed in the instrument. The only instrument to present the unidimensionality through the Rasch analysis, the (MAL)^{18,19}, included an inadequate number of participants¹⁴⁻¹⁶.

In terms of internal consistency, a sample of 7 participants is indicated for each item of the instrument, requiring a minimum of 100. For example, for an instrument of 30 items, a sample of 210 would be indicated¹⁴⁻¹⁶. It is recommended that internal consistency should be assessed in two ways: through the classic form, or by the item response theory. First, Cronbach's alpha should be calculated after the performance of the factorial analysis, which identifies the number of subscales where the alpha must be calculated¹⁴⁻¹⁶. Second, the Rasch mathematical model is indicated to assess the unidimensionality of the items, verifying the presence of items that can be adjusted or removed from the instrument¹⁴⁻¹⁶.

The flaws observed in the construct validity of the instruments^{20,25,26,28-30} generate uncertainties about the degree to which the Brazilian versions of the included instruments truly measure the correct construct. It is recommended that 100 participants be assessed and that hypotheses be previously formulated about the direction and magnitude of the expected correlation between the scores of the tested instruments and the comparator instrument¹⁴⁻¹⁶. The responsiveness and the criterion validity were not analyzed in any of the studies. The criterion validity is analyzed to verify the degree with which the scores of the instruments are an adequate reflection of the "gold-standard". However, for motor function measurement, no such instrument was observed in Brazil.

The absence of the responsiveness study, observed in all of the instruments, hampers the identification of the ability of these instruments to detect changes

in the assessed construct over time. Therefore, there is no evidence that it will be possible to quantify any motor function changes in post-stroke individuals in clinical research¹⁴⁻¹⁶.

Finally, the interpretability of the obtained scores in these instruments still has not been clarified. Despite the fact that the ceiling and floor effects in the Posture Assessment Scale and Trunk Deficiency Scale were analyzed and had favorable results, the sample size in both studies was inadequate. None verified the minimum important change (MIC) or the minimum important difference (MID). These results are relevant because the MIC is the smallest change in the construct score the patients observe as important and the MID corresponds to the minimum difference in the construct among patients that is considered important¹⁴⁻¹⁶. None of the instruments were tested for their interpretability and responsiveness. As such, it remains unknown whether these instruments are able of measuring clinical changes over time.

● Final considerations

Future studies should revise the cross-cultural adaptation processes, following all of the recommended stages (translation, synthesis of translations, back translation, expert committee, and pre-test). Moreover, the measurement properties should be analyzed with an adequate number of participants and the application of statistical methods that reflect the validity of each property. The results of this review point out that health professionals must be cautious when selecting instruments to assess post-stroke motor function for use in research and clinical practice in Brazil.

● References

1. Soriano FFS, Baraldi K. Escalas de avaliação funcional aplicáveis a pacientes pós- Acidente Vascular Encefálico. *Con Scientia e Saúde*. 2010;9(3):521-30.
2. Pomeroy VM, King LM, Pollock A, Alison B, Peter L. Electrostimulation for promoting recovery of movement or functional ability after stroke. *Cochrane Database Syst. Rev.* 2006(2):1-74. <http://dx.doi.org/10.1002/14651858.CD003241.pub2>.
3. Pollock A, Gill B, Pomeroy VM, Langhorne P. Physiotherapy treatment approaches for the recovery of postural control and lower limb function following stroke. *Cochrane Database Syst. Rev.* 2007(1):1-92. <http://dx.doi.org/10.1002/14651858.CD001920.pub2>.
4. Leitão A. Manual da Classificação Internacional da Funcionalidade Incapacidade e Saúde (CIF). Lisboa: OMS; 2004.
5. Falcão IV, Carvalho EMF, Barreto KML, Lessa FJD, Leite VMM. Acidente Vascular Cerebral precoce: implicações

- para adultos em idade produtiva atendidos pelo Sistema Único de Saúde. *Rev Bras Saude Mater Infant*. 2004;4(1):95-102. <http://dx.doi.org/10.1590/S1519-38292004000100009>.
6. Ferreira LTD, Pinto EB. Perfil Funcional dos pacientes acometidos por acidente vascular encefálico de um centro de referência em Salvador. *Fisioter Bras*. 2008;9(2):81-5.
7. Hendricks HT, van Limbeek J, Geurts AC, Zwarts MJ. Motor Recovery after Stroke: A systematic Review of the Literature. *Arch Phys Med Rehabil*. 2002;83(11):1629-37. <http://dx.doi.org/10.1053/apmr.2002.35473>. PMID:12422337.
8. Ceccato RB, Almeida CI. O planejamento da reabilitação na fase aguda após acidente vascular encefálico. *Acta Fisiátrica*. 2010;17(1):33-47.
9. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186-91. <http://dx.doi.org/10.1097/00007632-200012150-00014>. PMID:1124735.
10. Wild D, Grove A, Martin M, Eremenco 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(2):94-104. <http://dx.doi.org/10.1111/j.1524-4733.2005.04054.x>. PMID:15804318.
11. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42. <http://dx.doi.org/10.1016/j.jclinepi.2006.03.012>. PMID:17161752.
12. Kirshner B, Guyatt G. A Methodological framework for assessing health indices. *J Chronic Dis*. 1985;38(1):27-36. [http://dx.doi.org/10.1016/0021-9681\(85\)90005-0](http://dx.doi.org/10.1016/0021-9681(85)90005-0). PMID:3972947.
13. Terwee CB, Jansma EP, Riphagen II, De Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurements instruments. *Qual Life Res*. 2009;18(8):1115-23. <http://dx.doi.org/10.1007/s11136-009-9528-5>. PMID:19711195.
14. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. 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(4):539-49. <http://dx.doi.org/10.1007/s11136-010-9606-8>. PMID:20169472.
15. Mokink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;(10):22.
16. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology and definitions of measurements properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737-45. <http://dx.doi.org/10.1016/j.jclinepi.2010.02.006>. PMID:20494804.
17. Pereira ND, Michaelsen SM, Menezes IS, Ovando AC, Lima RC, Teixeira-Salmela LF. Reliability of the Brazilian version of the Wolf Motor Function Test in adults with hemiparesis. *Rev Bras Fisioter*. 2011;15(3):257-65. <http://dx.doi.org/10.1590/S1413-35552011000300013>. PMID:21829991.
18. Saliba VA, Magalhães LC, Faria CDCM, Laurentino GE, Cassiano JG, Teixeira-Salmela LF. Adaptação Transcultural e análise das propriedades psicométricas da versão brasileira do instrumento Motor Activity Log. *Rev Panam*

- Salud Publica. 2011;30(3):262-71. <http://dx.doi.org/10.1590/S1020-49892011000900011>. PMID:22069074.
19. Pereira ND, Ovando AC, Michaelsen SM, Anjos SM, Lima RCM, Nascimento LR, et al. Motor Activity Log-Brasil: reliability and relationships with motor impairments in individuals with chronic stroke. *Arq Neuropsiquiatr*. 2012;70(3):196-201. PMID:22392112.
 20. Michaelsen SM, Natalio MA, Silva AG, Pagnussat AS. Reliability of the translation and adaptation of the Test d'Évaluation des Membres Supérieurs des Personnes Âgées (TEMPA) to the Portuguese language and validation for adults with hemiparesis. *Rev Bras Fisioter*. 2008;12(6):511-9. <http://dx.doi.org/10.1590/S1413-35552008005000012>.
 21. Ferreira KN, Santos RL, Conforto AB. Psychometrics properties of the Portuguese version of the Jebsen-Taylor test for adults with mild hemiparesis. *Rev Bras Fisioter*. 2010;14(5):377-82. <http://dx.doi.org/10.1590/S1413-35552010005000018>. PMID:21203696.
 22. Pavan K, Cruz LC, Nunes MF, Menezes LG, Marangoni BE. Cross-cultural adaptation and validation of the Rivermead Mobility Index in stroke patients within the Brazilian cultural and language context. *Arq Neuropsiquiatr*. 2010;68(1):52-5. <http://dx.doi.org/10.1590/S0004-282X2010000100012>. PMID:20339653.
 23. Michaelsen SM, Rocha AS, Knabben RJ, Rodrigues LP, Fernandes CGC. Translation, adaptation and inter-rater reliability of the administration manual for the Fugl-Meyer assessment. *Rev Bras Fisioter*. 2011;15(1):80-8. <http://dx.doi.org/10.1590/S1413-35552011000100013>. PMID:21519719.
 24. Maki T, Quagliato EMAB, Cacho EWA, Paz LPS, Nascimento NH, Inoue MMEA, et al. Estudo de Confiabilidade da aplicação da escala de Fugl-Meyer no Brasil. *Rev Bras Fisioter*. 2006;10(2):177-83. <http://dx.doi.org/10.1590/S1413-35552006000200007>.
 25. Cincura C, Pontes-Neto OM, Neville IS, Mendes HF, Menezes DF, Mariano DC, et al. Validation of the National Institutes of health stroke scale, Modified Rankin Scale and Barthel Index in Brazil: The role of cultural adaptation and structured interviewing. *Cerebrovasc Dis*. 2009;27(2):119-22. <http://dx.doi.org/10.1159/000177918>. PMID:19039215.
 26. Caneda MAG, Fernandes JG, Almeida AG, Mugnol FE. Confiabilidade de escalas de comprometimento neurológico em pacientes com acidente vascular cerebral. *Arq Neuropsiquiatr*. 2006;64(3A):690-7. <http://dx.doi.org/10.1590/S0004-282X2006000400034>. PMID:17119821.
 27. Conte ALF, Ferrari PP, Carvalho TB, Relvas PCA, Neves RCM, Rosa SF. Reliability, comprehension and acceptability of the Portuguese version of the Motor Assessment Scale in stroke patients. *Rev Bras Fisioter*. 2009;13(5):405-11. <http://dx.doi.org/10.1590/S1413-35552009005000056>.
 28. Yonejama SM, Roiz RM, Oliveira TM, Oberg TG, Lima NMFV. Validação da versão brasileira da Escala de avaliação postural para pacientes após acidente vascular encefálico. *Acta Fisiatr*. 2008;15(2):96-100.
 29. Castelassi CS, Ribeiro EAF, Fonseca VC, Beinotti F, Oberg TD, Lima NMFV. Confiabilidade da versão brasileira da escala de deficiências de tronco em hemiparéticos. *Fisioter Mov*. 2009;22(2):189-99.
 30. Lima NMFV, Rodrigues SY, Fillipo TM, Oliveira R, Oberg TG, Cacho EWA. Versão brasileira da Escala de Comprometimento do tronco: um estudo de validade em sujeitos pós-acidente vascular encefálico. *Fisioter Pesqui*. 2008;16(3):248-53. <http://dx.doi.org/10.1590/S1809-29502008000300006>.
 31. Benaim C, Pérennou DA, Villy J, Rousseaux M, Pelissier JY. Validation of a standardized assessment of postural control in stroke patients: the Postural Assessment Scale for Stroke Patients (PASS). *Stroke*. 1999;30(9):1862-8. <http://dx.doi.org/10.1161/01.STR.30.9.1862>. PMID:10471437.
 32. Fujiwara T, Liu M, Tsuji T, Sonoda S, Mizuno K, Akaboshi K, et al. Development of a new measure to assess trunk impairment after stroke (Trunk Impairment Scale): its psychometric properties. *Am J Phys Med Rehabil*. 2004;83(9):681-8. <http://dx.doi.org/10.1097/01.PHM.0000137308.10562.20>. PMID:15314532.
 33. Verheyden G, Nieuwboer A, Mertin J, Preger R, Kiekens C, De Weerd W. The trunk impairment scale: a new tool to measure motor impairment of the trunk after stroke. *Clin Rehabil*. 2004;18(3):326-34. <http://dx.doi.org/10.1191/0269215504cr7330a>. PMID:15137564.
 34. Fugl-Meyer AR, Jaasko L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient: 1. A method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7(1):13-31. PMID:1135616.
 35. Jebsen RH, Taylor N, Trieschmann RB, Trotter MJ, Howard LA. An objective and standardized test of hand function. *Arch Phys Med Rehabil*. 1969;50(6):311-9. PMID:5788487.
 36. Carr JH, Shepherd RB, Nordholm L, Lynne D. Investigation of a new motor assessment scale for stroke patients. *Phys Ther*. 1985;65(2):175-80. PMID:3969398.
 37. Uswatte G, Taub E, Morris D, Light K, Thompson PA. The Motor Activity Log-28: assessing daily use of the hemiparetic arm after stroke. *Neurology*. 2006;67(7):1189-94. <http://dx.doi.org/10.1212/01.wnl.0000238164.90657.c2>. PMID:17030751.
 38. Brott TG, Adams HP Jr, Olinger CP, Marler JR, Barsan WG, Biller J, et al. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke*. 1989;20(7):864-70. <http://dx.doi.org/10.1161/01.STR.20.7.864>. PMID:2749846.
 39. Collen FM, Wade DT, Robb GF, Bradshaw CM. The Rivermead Mobility Index: a further development of the Rivermead Motor Assessment. *Int Disabil Stud*. 1991;13(2):50-4. <http://dx.doi.org/10.3109/03790799109166684>. PMID:1836787.
 40. Desrosiers J, Hébert R, Dutil E, Bravo G. Development and reliability of an upper extremity function test for the elderly: the TEMPA. *Can J Occup Ther*. 1993;60(1):9-16. <http://dx.doi.org/10.1177/000841749306000104>.
 41. Morris DM, Uswatte G, Crago JE, Cook EW 3rd, Taub E. The reliability of the wolf motor function test for assessing upper extremity function after stroke. *Arch Phys Med Rehabil*. 2001;82(6):750-5. <http://dx.doi.org/10.1053/apmr.2001.23183>. PMID:11387578.

Correspondence

Andrea Lemos

Universidade Federal de Pernambuco (UFPE)

Departamento de Fisioterapia

Avenida Professor Moraes Rego, 1235, Cidade Universitária

CEP 50670-901, Recife, PE, Brazil

e-mail: andreaemos4@gmail.com