ISSN 1413-3555
Rev Bras Fisioter, São Carlos, v. 15, n. 3, p. 257-65, May/June 2011

Reliability of the brazilian version of the Wolf Motor Function Test in adults with hemiparesis

Confiabilidade da versão brasileira do Wolf Motor Function Test em adultos com hemiparesia*

Natalia D. Pereira[1], Stella M. Michaelsen[1], Isabella S. Menezes[2], Angélica C. Ovando[1], Renata C. M. Lima[3], Luci F. Teixeira-Salmela[3]

Abstract

Background: The Wolf Motor Function Test (WMFT) evaluates the upper limb (UL) performance of adults with hemiparesis by combining time and quality of movement measures in both isolated movements and functional tasks. **Objectives:** To translate and adapt the WMFT form, functional ability scale (FAS) and manual to Brazilian Portuguese and evaluate the intra and inter-rater reliabilities. **Methods:** Fifteen individuals with a mean age of 57.9±11.1 years and a mean time since stroke onset of 68.5±53.5 months participated. The WMFT was administered by one physiotherapist based on information in the manual, and video observations were assessed by two other independent physical therapists. Information regarding compensatory movements was included in the FAS. Intra-class correlation coefficients (ICCs) and Bland-Altman plots were calculated to examine the intra- and inter-rater reliabilities for performance time and FAS, whereas weighted kappa (Kp) was used to examine the agreement strength for FAS. **Results:** The inter-rater ICC values for performance time were above 0.75 in 13 of the 15 tasks. For the FAS, they ranged from 0.87-0.99 for all evaluated tasks, with Kp values ranging from 0.63-0.92. For intra-rater reliability, the ICC ranged from 0.99-1.0 and from 0.96-1.0 for time measurement and FAS, respectively. Kp values ranged from 0.79-0.96 for individual and 0.93 for total scores. **Conclusion:** The Brazilian version of the WMFT showed adequate intra- and inter-rater reliabilities for evaluating the paretic UL of individuals with stroke.

Keywords: translating; stroke; hemiplegia; upper extremity; Wolf Motor Function Test; reproducibility of results.

Resumo

Contextualização: O *Wolf Motor Function Test* (WMFT) avalia o membro superior (MS) de adultos com hemiparesia combinando medidas de tempo e qualidade de movimento em movimentos isolados e em tarefas funcionais. **Objetivos:** Traduzir e adaptar para a língua portuguesa o formulário, a escala de habilidade funcional (EHF) e o manual de aplicação do WMFT e avaliar a confiabilidade intra e inter-observadores. **Métodos:** Participaram 15 indivíduos com média de idade de 57,9±11,1 anos e 68,5±53,5 meses pós acidente vascular encefálico (AVE). O WMFT foi aplicado por um fisioterapeuta utilizando as informações do manual e cotado por dois outros fisioterapeutas independentes pela observação dos vídeos. Foram acrescentadas informações mais detalhadas na EHF sobre a movimentação compensatória em relação à escala original. A confiabilidade intra e interobservadores do desempenho no tempo e da EHF dos itens individuais e do escore total foi avaliada pelo Coeficiente de Correlação Intraclasse (CCI) e pelo método Bland e Altman. Kappa ponderado (Kp) foi utilizado para avaliar a concordância intra e interobservadores da EHF. **Resultados:** O CCI interobservador do desempenho no tempo foi >0,75 em 13 das 15 tarefas. A EHF apresentou CCI interobservador entre 0,87 e 0,99 em todas as tarefas e Kp entre 0,63 e 0,92. O CCI intraobservador do tempo variou entre 0,99 e 1 e na EHF, entre 0,96 e 1. O Kp intraobservador na EHF nas tarefas variou entre 0,79 e 0,96, sendo 0,93 para o escore total. **Conclusão:** A versão brasileira do WMFT demonstrou confiabilidade adequada para avaliar o MS parético pós-AVE.

Palavras-chave: tradução; acidente cerebral vascular; hemiplegia/hemiparesia; extremidade superior; Wolf Motor Function Test; reprodutibilidade dos testes.

Received: 04/11/2010 – **Revised:** 04/04/2011 – **Accepted:** 15/04/2011

[1] Post Graduation Program in Human Movement Sciences, Universidade do Estado de Santa Catarina (UDESC), Florianópolis, SC, Brazil
[2] Association of Assistance to Disabled Children (AACD), São Paulo, SP, Brazil
[3] Post Graduation Program in Rehabilitation Science, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil
*This study was presented at the 1st Brazilian Congress of Neurofunctional Physical Therapy, Petrópolis, RJ.
Correspondence to: Natalia Duarte Pereira, PPG Ciências do Movimento Humano - CEFID/UDESC, Laboratório de Controle Motor – LADECOM, Rua Pascoal Simone, 358, Coqueiros, CEP 88080-350, Florianópolis, SC, Brasil, e-mail: nat_duarte@yahoo.com.br

Introduction

It is estimated that 45 to 75% of adults who have suffered a stroke have difficulty using the hemiparetic upper limb (UL) in activities of daily living (ADLs) in the chronic stage[1]. Reliable systematic reviews provide adequate information for clinical decision making about UL rehabilitation and demonstrate the effects of an intervention designed to improve the functionality, coordination and precision of UL movements[2,3].

The Wolf Motor Function Test (WMFT) was initially developed to evaluate the effects of constraint-induced movement therapy in individuals with hemiparesis[1-3]. The original version consisted of 21 tasks organized according to the joints involved (from the shoulder to the fingers) and the difficulty level (from gross to fine motor activity), evaluating the UL function by means of one or multiple joint movements and functional tasks. As shown in Appendix 1, this test was subsequently modified to a version with 17 sequential tasks[4] in order to simplify its application. The WMFT evaluates the speed of task performance, quantifies the quality of movement by means of a functional ability scale (FAS) and measures handgrip strength and shoulder flexion strength in two specific tasks[2]. The final result provides the average task completion time, the median FAS score, handgrip strength in kgf and the weight lifted in grams in the shoulder flexion movement. Because they involve different units, the strength tasks are not included in the final performance time or the FAS[3]. Handgrip strength measurement by dynamometer has already been tested for reliability[5], and thus the two strength measures were not included in this reliability study.

The WMFT tasks must be videotaped from a standardized position and distance, and scoring is carried out by analyzing the videos. The intraclass correlation coefficient (ICC) of the inter-rater reliability of the total FAS score in the original version of the WMFT was 0.88. The measurement of task performance speed showed high reliability both in direct measurement[6] (ICC=0.97) and in video scoring (ICC=0.96)[5]. For the results to be reliable, reading the manual and using the specified material are recommended[6].

The FAS has six levels, with "zero" indicating no attempt to move the UL that is being tested and "five" indicating that the movement seems to be normal. In the original version, subjects are scored according to the presence of abnormal movements such as lack of coordination between the segments, lack of fine coordination, fluidity or precision, slow task performance and difficulty in performing resistance activities. Although other tests evaluating the paretic UL in patients with hemiparesis have already been translated and adapted to Portuguese, the WMFT is the only one that combines measurements of time and quality of movement performance in both isolated movements of specific joints and complex functional tasks; thus, it is an evaluation that can be applied to patients with different levels of impairment. Among widely used tests that have been validated in Brazil, the Test d'Évaluation des Membres Supérieurs de Personnes Âgées (TEMPA)[7] evaluates only the function component, whereas the Fugl-Meyer scale[8] evaluates only movement.

The TEMPA[7] evaluates UL focal function according to time, functional score and observational analysis of task performance. It includes four unilateral tasks and four bilateral tasks that represent ADLs and is recommended when the aim is to evaluate activity limitations. The Fugl-Meyer scale[8] evaluates the recovery of the post-stroke paretic UL by analyzing movement components according to the presence/absence of abnormal synergies. Fugl-Meyer is recommended for the evaluation of impairments in body function or structure, without considering functionality. However, the WMFT can offer the advantage of evaluating both the function and body structure components[9]. Nevertheless, Massie et al.[10] reported that in the outcome measures used in constraint-induced movement therapy (CIMT), including the WMFT, aspects of movement quality related to compensatory movements and incoordination between the segments, which are characteristic of this population, are not considered[11].

Thus, the aims of the present study were to translate and adapt the WMFT form and application manual to Brazilian Portuguese, adding information to the FAS regarding compensatory movements, and to evaluate the test's intra- and inter-rater reliability using video.

Methods

Translation and adaptation of the WMFT

The WMFT was created in 1989[2] in American English and the application manual was developed in 2000[4]. In the present study, the form, manual and FAS were independently translated into Portuguese by two Brazilian physical therapists and a Brazilian occupational therapist. These professionals, fluent in English, were trained by the group of authors of the latest version of the test at the University of Alabama at Birmingham (UAB), USA. The translated manual is available on request from the authors.

The manual was translated with the consent of the original author. Information was added to the FAS (Table 1A)

regarding the original scale (Table 1B) including details about compensatory movement, incoordination between the segments and further explanation added about the movement quality items, which are based on the Quality of Movement Scale used in CIMT[12].

This version was back-translated by a physical therapist fluent in both languages and then sent to the UAB group for approval. The group considered that the original concepts were preserved in the adapted version. A group of six physical therapists experienced in neurological rehabilitation considered the adapted version sufficiently clear[13].

Before scoring the videos, the untrained examiner read the manual and discussed problematic issues with the trained examiner. Questions were also discussed with researchers from UAB, who sent updates of the manual to facilitate scoring. Clarifications included: if in doubt between two levels, the lower score should be adopted; in case of a discrepancy between the quality of shoulder movement and hand or elbow movement, the task category must be taken into consideration. For example, pinching is the tested movement in the paper clip lifting task, so hand performance is of primary importance, even if the quality of shoulder movement is lower. Thus, information of this type was added to the manual.

Application of the WMFT

The WMFT was applied by a physical therapist who had no previous experience with the test and relied completely on the information contained in the manual. While the test was being applied, the 15 tasks were videotaped according to the guidelines in the manual. The videos were subsequently evaluated by two other physical therapists. The two strength tasks were not performed (7 and 14), as was the procedure in the original reliability studies[1,3]. The physical therapist who applied the test did not participate in the scoring. The examiner directed each task according to the instructions and commands described in the manual. Each task was demonstrated both slowly and quickly to the subject, who was then requested to do it as quickly as possible.

To evaluate performance speed, the video was used to time the tasks from the moment of the verbal command "Go!" until the moment it was completed. The six-level FAS was used to evaluate the quality of movement in each task (Table 1A). The total score for performance time was calculated by the mean performance time of all tasks. When the individual was unable to accomplish a certain task, a score of 121 seconds was given, since 120 seconds is the maximum time allowed for each task[14]. In this study, the mean and median of the score were used, since using only the mean can lead to arbitrary representation of the values when incomplete tasks are involved. Functional performance was evaluated by the median of the scores of all tasks[15].

The WMFT was applied using a model that establishes the standard position of each object and its initial and/or final position in the tasks. All tasks were videotaped and FAS was evaluated according to the instructions in the manual.

Table 1. Modified Functional Ability Scale.

Functional Ability Scale (FAS)	
A - Modified version	B – Literal translation
0- Does not attempt with upper extremity (UE) being tested.	0 – Does not attempt with upper extremity (UE) being tested
1- Despite attempt to use the upper limb, the limb being tested is not functional because it does not complete the task. The UE which is not being tested can be used to move the UE being tested. The movement observed has predominantly abnormal synergy and incoordination between limb segments.	1- UE being tested does not participate functionally; however, attempt is made to use the UE. In unilateral tasks the UE not being tested may be used to move the UE being tested.
2- Does, but requires assistance of the evaluator or the UE not being tested for minor adjustments or changes in position, or requires more than two attempts to complete the task or performs it very slowly. The movement is influenced by abnormal synergy or the movement is performed with excessive trunk's, head's or contralateral UE compensatory movements or lack of proximal control or fine motor skill. In bilateral tasks the UE being tested may serve only as a helper.	2- Does, but requires assistance of the UE not being tested for minor readjustments or change position, or requires more than two attempts to complete, or accomplishes very slowly. In bilateral tasks the UE being tested may serve only as a helper.
3- Does, but movement is influenced by some degree of abnormal synergy or compensation, or primitive grasp patterns. The movement is performed slowly or with abnormal effort; moderate incoordination and lack of precision.	3- Does, but movement is influenced to some degree by synergy or is performed slowly or with effort.
4- Does; movement is close to normal *, but it is slightly slow; may lack precision, fine coordination, or movement fluidity	4- Does, movement is close to normal*, but slightly slower, may lack precision, fine coordination or fluidity.
5- Does; normal movement* Fluid and coordinated activity; movement velocity appears within normal limits.	5- Does; movement appears to be normal*.

(*)In order to determine patterns of normality, the less affected upper limb can be used for comparison. However, upper limb dominance must be taken into consideration.

Table 2. Clinical characteristics of the participants.

Participant	Age (years)	Sex	Chronicity (months)	Fugl-Meyer (UE score)	Mini Mental (score)	Affected side	Dominance
1	46	M	91	50	28	R	R
2	60	M	96	46	22	R	R
3	76	F	129	43	28	R	R
4	68	M	8	58	28	L	R
5	55	M	41	52	29	L	R
6	62	M	173	49	26	L	R
7	53	F	124	52	28	R	R
8	55	F	60	53	30	R	R
9	72	M	8	10	26	L	R
10	58	F	84	58	30	L	R
11	42	M	18	10	22	R	R
12	48	M	9	28	26	L	R
13	40	M	10	47	28	L	R
14	62	M	54	8	27	L	R
15	72	F	123	60	22	R	R
Average (SD)	57.9 (11.1)	10M/5F	68.5 (53.5)	41.6 (18.3)	26.7 (2.7)	8L/7R	15R

UE=Upper Extremity; M=Male; F=Female: R=Right; L=Left; SD=Standard Deviation.

Participants

The study included a convenience sample of 15 adults (57.9±11.1 years) with chronic hemiparesis (68.5±53.5 months after stroke) who were recruited from the Clinical School of Physical Therapy of the Universidade do Estado de Santa Catarina (UDESC), Florianópolis, SC, Brazil. After signing the consent form approved by the UDESC Research Ethics Committee (Nº 144/2009), the Fugl-Meyer motor recovery assessment[8] was applied and all individuals who met the following inclusion criteria were included: age >18 years, being at least six months post-stroke and having no cognitive impairment according to the Mini Mental State Examination (MMSE) with the cutoff points recommended by Brucki et al.[16]. Individuals with other neurological diseases, those who presented full motor recovery and those with a totally paralyzed limb were excluded.

Intra and inter-rater reliability

Based on the videos, test performance was evaluated by two independent examiners who, without conferring with each other, recorded the task performance time and scored the FAS. To determine intra-rater reliability, the same examiners reevaluated the videos after one month.

Statistical analysis

ICCs with confidence intervals (CI) of 95% were calculated to evaluate intra- and inter-rater reliability for time measurement and FAS[17]. Absolute agreement ICC was also used, which determines the correlation of evaluations in terms of absolute scores[18].

Weighted kappa (Kp) was used to evaluate intra- and inter-rater reliability of the FAS since it is recommended for evaluating ordinal scales[19]. However, the ICC was also used for comparison with the literature. For a more detailed analysis of scoring differences by both examiners, a Bland-Altman plot was used to compare the total time and FAS scores[20]. The level of significance was set at 5%.

Results

Sample characterization

The sample consisted of individuals with mild UL (MS≥50) impairment (46%), moderate impairment (30<MS<49) (27%) and severe impairment (MS≤30) (27%)[8] (Table 2).

Inter-rater reliability

Table 3 describes the mean and median of time, median FAS, inter-rater ICC, and the FAS Kp value for each task. For time measurement, the inter-rater ICC ranged from 0.87-1 in 13 of 15 tasks[21], and the tasks "forearm to table" and "hand to box" showed lower than adequate ICC values. When all tasks were considered (total score), the ICC values were 0.94 (0.82-0.98) for the FAS and 0.99 (0.98-0.99) for time; the inter-rater absolute agreement was 0.94 (0.83-0.98) for the FAS and 0.99 (0.98-0.99) for total time score.

Table 3. Descriptive statistics, inter-rater reliability and upper and lower limit of the 95% confidence intervals for the individual items of the WMFT.

Time (seconds)	Rater 1 Median (min-max)	Rater 2 Median (min-max)	Rater 1 Mean (SD)	Rater 2 Mean (SD)	ICC (95% IC)
Forearm to table	0.9 (0.6-4.6)	1.1 (0.5-121)	1.4 (1.2)	9.3 (30.9)	NS
Forearm to Box	1.2 (0.5-5.9)	1.3 (0.5-4.6)	1.5 (1.4)	1.63 (1.1)	0.96 (0.88-0.99)**
Extend elbow	1.4 (0.6-121)	1.9 (0.5-121)	10.1 (30. 7)	10.3 (30.7)	1 (1-1)**
Extend elbow	0.9 (0.4-121)	1 (0.2-121)	10.5 (31.2)	10.8 (31.2)	1 (1-1)**
Hand to table	0.8 (0.4-6.8)	0.9 (0.5-7.2)	1.4 (1.6)	1.5 (1.7)	0.99 (0.98-0.99)**
Hand to Box	1.1 (0.4-121)	1.5 (0.5-121)	9.9 (30.8)	25.5 (49.4)	0.67 (0.29-0.89) *
Reach and retrieve	1.1 (0.4-121)	1.4 (0.3-121)	3 (1.6)	3 (1.5)	1 (1-1) **
Lift can	2.8 (1.8-121)	2.8 (1.8-121)	2.4 (1.2)	2.5 (1.3)	0.87 (0.65-95)**
Lift pencil	6.3 (1.4-121)	6.3 (1.5-121)	46.7 (55.5)	37.3 (48.3)	0.93 (0.78-0.97)**
Lift paper clip	7.9 (1.1-121)	7.7 (1.3-121)	43.4 (56.9)	36.2 (53.1)	0.93 (0.79-0.98)**
Stack checkers	11.5 (4-121)	12 (4.1-121)	48 (53.9)	48 (53.9)	1 (1-1)**
Flip cards	23.8 (8-121)	24 (7.8-121)	52.1 (51.8)	53.4 (51.1)	99 (0.99-0.99)**
Turn key in lock	17.2 (4.6-121)	17.8 (5.4-121)	47.9 (53.8)	48.1 (53.6)	1 (1-1)**
Fold towel	15.3 (5.4-121)	15.4 (5.3-121)	29.9 (38.5)	29.8 (38.6)	1 (1-1)**
Lift basket	4.9 (2.8-121)	5.6 (2.7-121)	28.8 (47.9)	29.1 (47.8)	1 (1-1)**
FAZ	Median (min-max)	Median (min-max)	Kappa Value (95% IC)		ICC (95% IC)
Forearm to table	3 (1-5)	3 (1-5)	0.83 (0.66-1)		0.97 (0.89-0.98)**
Forearm to Box	2 (2-5)	2 (2-5)	0.81 (0.6-1)		0.94 (0.83-0.98) **
Extend elbow	3 (1-5)	3 (1-5)	0.71 (0.44-0.97)		0.92 (0.75-0.97)**
Extend elbow	3 (1-5)	3 (1-5)	0.63 (0.37-0.89)		0.9 (0.7-0.97)*
Hand to table	3 (2-5)	3 (2-5)	0.62 (0.34-0.9)		0.88 (0.63-0.96)**
Hand to Box	4 (1-5)	4 (1-5)	0.73 (0.5-0.95)		0.93 (0.81-0.98)**
Reach and retrieve	3 (0-5)	3 (1-5)	0.92 (0.83-1)		0.99 (0.96-0.99)**
Lift can	4 (1-4)	4 (1-5)	0.8 (0.59-1)		0.93 (0.79-0.98)**
Lift pencil	1 (1-4)	1 (1-5)	0.72 (0.55-0.89)		0.93 (0.8-0.98)**
Lift paper clip	3 (0-4)	3 (0-4)	0.84 (0.69-0.99)		0.97 (0.9-0.99)**
Stack checkers	1 (1-4)	1 (1-4)	0.78 (0.56-1)		0.93 (0.78-0.98)**
Flip cards	1 (1-3)	1 (1-4)	0.67 (0.41-0.93)		0.9 (0.7-0.97)**
Turn key in lock	2 (1-4)	2 (1-4)	0.77 (0.56-0.97)		0.93 (0.81-0.98)**
Fold towel	1 (0-4)	1 (0-4)	0.77 (0.6-0.95)		0.95 (0.84-0.98)**
Lift basket	4 (1-4)	4 (1-4)	0.73 (0.5-0.96)		0.87 (0.65-0.95)**

FAS=Functional ability scale; sd=standard deviation; (min-max) =minimal and maximal values ICC=intraclass correlation coefficient; CI= confidence interval; *<0.05; **<0.01; NS=not significant.

The Bland-Altman plot demonstrated inter-rater reliability. For the FAS, the average difference between the two evaluations did not differ significantly from zero and the limits of agreement represented 8.2 and 9.6% of the scale range. Similarly, the inter-rater difference for time was very close to zero.

Intra-rater reliability

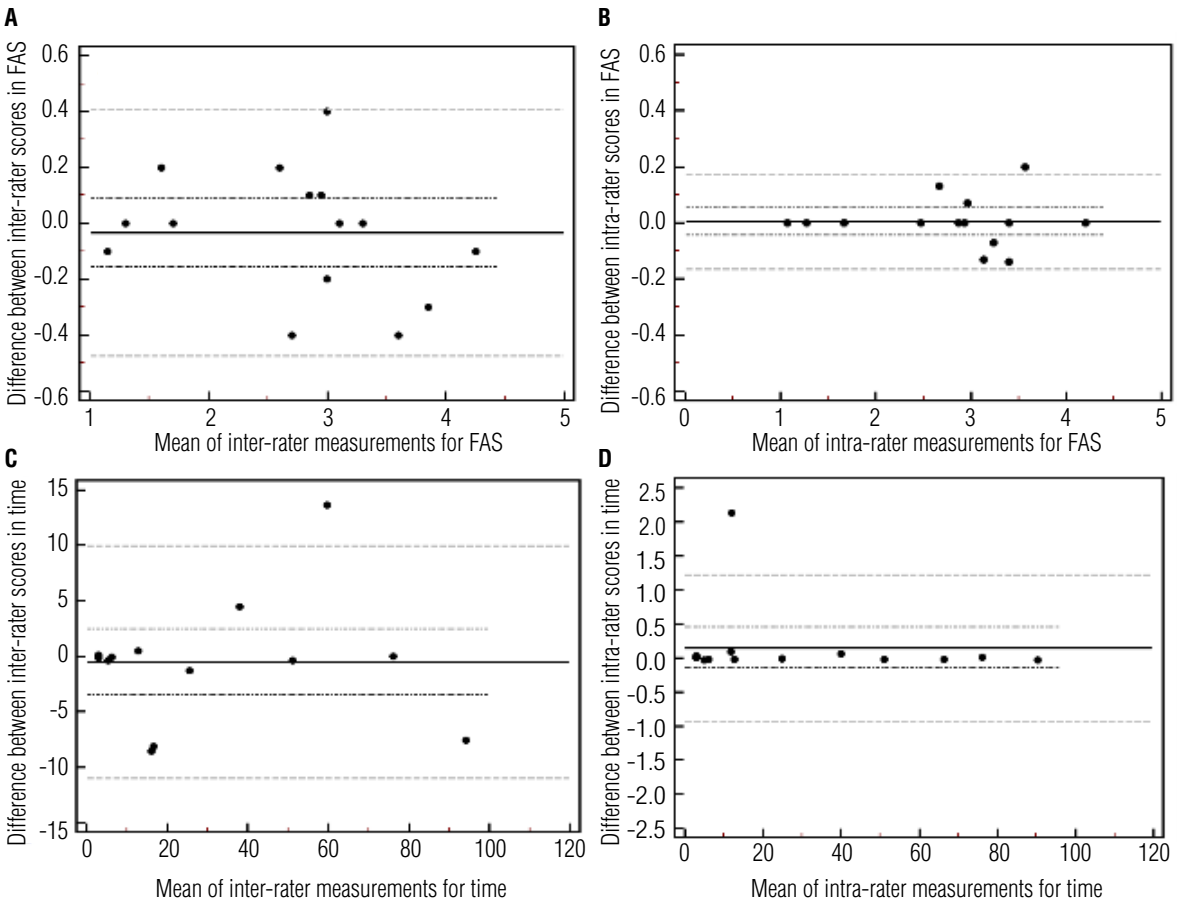
The ICC values for time measurement, with the exception of the task "forearm to box", ranged from 0.99-1 and for the FAS from 0.96-1. For the FAS, the Kp value ranged from 0.79 ("lift the basket") to 0.96 ("reach and retrieve"), with Kp=0.93 (CI=0.78 to 1.0) for the total score.

The Bland-Altman plot showed that the average difference between the two evaluations did not differ significantly from zero, the limits of agreement represented 3.4% of the FAS range, and the time difference was very close to zero (Figure 1).

Discussion

This study translated and adapted the WMFT to Brazilian Portuguese and determined its intra- and inter-rater reliability for individuals with hemiparesis whose impairment ranged from mild to severe. In general, the results were similar to previous reliability studies[4,15].

The reliability of inter-rater time measurement was assessed to determine if the two examiners followed the same criteria for gauging the beginning and end of the task as well as failure to complete the task. The inter-rater reliability of



A= Inter-rater functional ability scale (FAS) scores; B=intra-rater FAS scores; C= inter-rater time scores; D=intra-rater time scores. The middle horizontal continuous black lines indicate the mean differences and the confidence intervals are presented in black dotted line. The gray dotted lines (more distant) indicate the upper and lower limits of agreement.

Figure 1. Scatter-plots of the differences between the two WMFT measurements and mean of total individual scores.

time measurement in the translated version of the WMFT was adequate, with an ICC similar to that found by Morris et al.[4] in the original version.

When individually analyzed, 13 of the 15 tasks had high levels of inter-rater reliability for time measurement. The tasks "forearm to table" and "hand to box" showed unsatisfactory and moderate levels (ICC<0.75), respectively[21]. This was due to the fact that one of the examiners considered the task "forearm to table" incomplete for two subjects and the task "hand to box" incomplete for another subject, which resulted in scores of 121 seconds, while the other examiner considered the task completed and recorded a different total time. However, one of the videos was stopped immediately after task completion, which made it difficult to judge the final position and whether the task was completed or not. This was considered a limitation of the study. Such a discrepancy could have been minimized by continuing to record for a few seconds after task completion. This information was added to the manual in the instructions about videotaping.

The median can characterize the performance of time measurement with greater reliability since it is less influenced by outlier values from unfinished tasks (121 seconds)[22]. However, for evaluating the effect of an intervention, the mean may be more sensitive both in time and in FAS; for example, if the subject is able to perform one of the tasks that he did not perform before an intervention, this may be best represented by the mean[15].

The absolute agreement between evaluators in the translated and adapted version of the FAS was adequate (ICC=0.94) and slightly higher than values found in the original version[4] (ICC=0.88). These results suggest that the examiners not only scored functional ability in a similar manner but also tended to give to the same individuals the same absolute scores. Regarding individual tasks in the FAS, all showed adequate inter-rater reliability (ICC>0.87) with good Kp values (Kp>0.61)[17].

The intra-rater ICC value for total score of time measurement and FAS was 100%, despite slight differences observed in the mean, median and ICC values of the individual

tasks. The absolute agreement for the FAS was more sensitive to these small changes (Kp=0.93). To evaluate the FAS, Kp is recommended since it evaluates ordinal scales and is more rigid regarding a scoring difference of two or more levels between two examiners, i.e., if one examiner gives a score of "2" and the other gives a "4". All other tasks would show an agreement level between "good" and "very good" in Kp, despite showing lower values than the ICC[17]. Although Kp is the most suitable method for ordinal scales such as the FAS, the ICC was also used for comparison with the literature.

Since this study's results depended on the evaluation of a video by the same examiner one month apart, this may have resulted in a higher ICC than that of the original version (0.90 for time and 0.95 for FAS), in which the same examiner administered two tests two weeks apart[4]. The reason for evaluating the same video was to eliminate any variability due to the subject in two test applications so as to better consider the examiner's ability to judge the beginning and end of the same video after a time interval between the two observations. These high reliability coefficients are probably due to the detailed scoring description provided in the manual, the degree to which the test was standardized and the fact that the Brazilian version of the FAS is more detailed. Sanford et al.[23] highlighted the importance of standardizing the manuals of assessment instruments in order to reduce measurement errors.

The WMFT includes a variety of movements that can be useful in both the clinical evaluation of patients with hemiparesis and as a research tool[24,25]. It is a scale that evaluates motor deficits by means of quantitative variables, performance time, fine coordination and fluidity, among other clinically relevant characteristics[2]. For example, compensatory movements of the trunk have been reported to limit UL motor recovery after stroke[26]. The modified version of FAS explicitly demonstrates how to distinguish these movements and, thus, may assist in identifying arm movement recovery as opposed to the use of compensatory strategies which would thus help the clinician decide on a therapeutic approach. Like the TEMPA, the WMFT includes a qualitative evaluation of movement, an advantage not shared by tests such as Jebsen Hand Function Test[24], which scores only the speed of task performance. Although more difficult to objectively quantify, these characteristics are important for evaluating movement. For the quality score to be reliable, it is suggested that the videos should be evaluated by examiners who understand the main compensation strategies of patients with different levels of impairment.

In the Brazilian version of the FAS, further details on compensatory movement, abnormal synergies, proximal control and primitive grip patterns were added. This information was based on the Quality of Movement Scale used in the CIMT, which was created by the same group responsible for the WMFT. In fact, the WMFT was designed to evaluate the results of this therapeutic intervention[4,12,25]. The authors of the WMFT agreed to allow the original concept of the scale to remain in the Brazilian Portuguese version. This modification aimed to take into account changes in movement that are commonly observed in hemiparetic subjects and that were not clearly covered in the scale. Another reason for the modification was to allow better classification of patients in each of the six levels.

After the time measurement and FAS evaluation, the examiners discussed the points that were difficult to interpret in the manual, which were then rewritten. For example, the initial position in the eighth task was redescribed with greater clarity ("the forearm in neutral position"). In addition to the instructions sent by UAB, a suggestion made by Whitall et al.[15] was added about instructing the patient not to use the less affected arm except in the bimanual task of folding the towel. Other modifications included some camera angles considered difficult to visualize for the functional evaluation, which were changed to a more favorable position (frontal) in the following tasks: lifting a can, lifting a pencil, lifting a paper clip, stacking checkers, flipping cards and folding a towel. It is essential that the videotaping meet the specifications of the manual regarding the position and distance of the camera in order to properly view the compensatory movements. Videotaping the entire body on the paretic side allows visualization of compensation due to anterior trunk displacement[3,26], while videotaping frontally and at close range facilitates the observation of compensatory movements made while gripping objects[11].

All subjects were able to perform at least the initial tasks of the test, even the most affected. A patient with a score of eight on the Fugl-Meyer scale was able to perform eight of the 15 tasks. The organization of the test regarding task complexity, i.e., from the easiest to the most complicated, motivated patients with greater difficulties. Other advantages of this test are that it also includes a bimanual task and uses common materials, unlike the Action Research Arm Test[6,27,28], which uses cylinders and wood blocks, materials with low ecological validity because they are not encountered in daily living[29,30]. Although specific equipment is required for the administration of the test, most of the items are common, inexpensive and easily obtained. This fact, combined with adequate reliability and internal consistency[6], makes the WMFT an excellent test for research purposes. One obstacle in clinical practice may be the time it takes to apply the test, which is about 30 minutes. However, in situations where it is important to document changes during treatment, the WMFT can provide valuable information.

The results indicate that the Brazilian version of the WMFT has appropriate intra- and inter-rater reliability for the total test score and that it can be used to evaluate functionality of the paretic UL of adults with hemiparesis. The video observation method, carried out according to the manual's instructions, proved to be a reliable means for evaluating the time and quality of movement, making it easier to apply the WMFT in clinical and research environments.

References

1. Morris DM, Taub E. Constraint-induced therapy approach to restoring function after neurological injury. Top Stroke Rehabil. 2001;8(3):16-30.
2. Wolf SL, Lecraw DE, Barton LA, Jann BB. Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients. Exp Neurol. 1989;104(2):125-32.
3. Winstein CJ, Miller JP, Blanton S, Taub E, Uswatte G, Morris D, et al. Methods for a multisite randomized trial to investigate the effect of constraint-induced movement therapy in improving upper extremity function among adults recovering from a cerebrovascular stroke. Neurorehabil Neural Repair. 2003;17(3):137-52.
4. 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.
5. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. J Hand Surg Am. 1984;9(2):222-6.
6. Wolf SL, Catlin PA, Ellis M, Archer AL, Morgan B, Piacentino A. Assessing Wolf motor function test as outcome measure for research in patients after stroke. Stroke. 2001;32(7):1635-9.
7. Michaelsen SM, Natalio MA, Silva AG, Pagnussat AS. Confiabilidade da tradução e adaptação do Test d'Évaluation des Membres Supérieurs de Personnes Âgées (TEMPA) para o português e validação para adultos com hemiparesia. Rev Bras Fisioter. 2008;12(6):511-9.
8. Michaelsen SM, Rocha AS, Knabben RJ, Rodrigues LP, Fernandes CGC. Tradução, adaptação e confiabilidade inter-examinadores do manual de administração da escala de Fugl-Meyer. Rev Bras Fisioter. 2011;15(1):80-8.
9. Farias N, Buchalla CM. A classificação internacional de funcionalidade, incapacidade e saúde da organização mundial da saúde: conceitos, usos e perspectivas. Rev Bras Epidemiol. 2005;8(2):187-93.
10. Massie C, Malcolm MP, Greene D, Thaut M. The effects of constraint-induced therapy on kinematic outcomes and compensatory movement patterns: an exploratory study. Arch Phys Med Rehabil. 2009;90(4):571-9.
11. Michaelsen SM, Jacobs S, Roby-Brami A, Levin MF. Compensation for distal impairments of grasping in adults with hemiparesis. Exp Brain Res. 2004;157(2):162-73.
12. Taub E, Uswatte G, King DK, Morris D, Crago JE, Chatterjee A. A placebo-controlled trial of constraint-induced movement therapy for upper extremity after stroke. Stroke. 2006;37(4):1045-9.
13. Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil. 2000;81(12 Suppl 2):S15-20.
14. Wolf SL, Thompson PA, Morris DM, Rose DK, Winstein CJ, Taub E, et al. The EXCITE trial: attributes of the Wolf Motor Function Test in patients with subacute stroke. Neurorehabil Neural Repair. 2005;19(3):194-205.
15. Whitall J, Savin DN Jr, Harris-Love M, Waller SM. Psychometric properties of a modified Wolf Motor Function test for people with mild and moderate upper-extremity hemiparesis. Arch Phys Med Rehabil. 2006;87(5):656-60.
16. Brucki SMD, Nitrini R, Caramelli P, Bertolucci PHF, Okamoto IH, Ivan H. Sugestões para o uso do MiniExame do Estado Mental no Brasil. Arq Neuropsiquiatr. 2003;61(3B):777-81.
17. Altman DG, editor. Practical statistics for medical research. London: Chapman & Hall; 1991.
18. Ottenbacher KJ, Taylor ET, Msall ME, Braun S, Lane SJ, Granger CV, et al. The stability and equivalence reliability of the functional independence measure for children (WeeFIM). Dev Med Child Neurol. 1996;38(10):907-16.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics.1977;33(1):159-74.
20. van der Lee JH, Beckerman H, Knol DL, de Vet HC, Bouter LM. Clinimetric properties of the motor activity log for the assessment of arm use in hemiparetic patients. Stroke. 2004;35(6):1410-4.
21. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. 3rd ed. New York: John Wiley & Sons; 1981.
22. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. J Hand Surger Am. 1984;9(2):222-6.
23. Sanford J, Moreland J, Swanson LR, Stratford PW, Gowland C. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. Phys Ther. 1993;73(7):447-54.
24. Ferreiro KN, dos Santos RL, Conforto AB. Psychometric properties of the Portuguese version of the Jebsen-Taylor test for adults with mild hemiparesis. Rev Bras Fisioter. 2010;14(5):377-82.
25. Taub E, Uswatt G. Constraint-induced movement therapy: answers and questions after two decades of research. NeuroRehabilitation. 2006;21(2):93-5.
26. Michaelsen SM, Dannenbaum R, Levin MF. Task-specific training with trunk restraint on arm recovery in stroke: randomized control trial. Stroke. 2006;37(1):186-92.
27. Paz LPS, Borges G. Teste da ação da extremidade superior como medida de comprometimento após AVC. Rev Neurocienc. 2007;15(4):277-83.
28. Hsieh CL, Hsueh IP, Chiang FM, Lin PH. Inter-rater reliability and validity of the action research arm test in stroke patients. Age Ageing. 1998;27(2):107-13.
29. Newell KM. Motor skill acquisition. Annu Rev Psychol. 1991;42:213-37.
30. Newell KM. On task and theory specificity. J Mot Behav. 1989;21(1):92-6.

Appendix 1. WMFT - formulário de coleta de dados (Portuguese version).

Nome do paciente: _____

Data:____/____/____

Teste (checagem 1): Pré-tratamento: ____ Pós-tratamento:____ Seguimento:____

Teste do braço (checagem 1): Mais afetado_____ Menos afetado_____

Tarefa	Tempo	Habilidade funcional (HF)	Comentário	
1. Antebraço na mesa				012345
2. Antebraço na caixa				012345
3. Extensão de cotovelo				012345
4. Extensão do cotovelo (com peso)				012345
5. Mão na mesa				012345
6. Mão na caixa				012345
7. Com peso na caixa*				_____g
8. Alcançar e retroceder				012345
9. Levantar lata				012345
10. Levantar lápis				012345
11. Levantar clipe de papel				012345
12. Empilhar peças				012345
13. Virar cartas				012345
14. Força de preensão*				_____Kgf
15. Virar chave				012345
16. Dobrar toalha				012345
17. Levantar cesta				012345

* os itens de força não são incluídos no desempenho final do tempo ou na HF

Descrição das tarefas do WMFT

1. Antebraço na mesa (de lado): colocar o antebraço na mesa fazendo abdução de ombro.
2. Antebraço na caixa (de lado): colocar o antebraço na caixa fazendo abdução de ombro.
3. Extensão de cotovelo (de lado): Levar a mão do outro lado da mesa estendendo o cotovelo.
4. Extensão de cotovelo com peso (de lado): Empurrar o peso para o outro lado da mesa estendendo o cotovelo.
5. Mão na mesa (de frente): Colocar a mão testada na mesa.
6. Mão na caixa (de frente): Colocar a mão na caixa.
8. Alcançar e retroceder (de frente): Puxar peso de 1 kg através da mesa usando flexão de cotovelo, antebraço na posição neutra e mão em concha.
9. Levantar lata (de frente): Levantar a lata e aproximá-la dos lábios com preensão cilíndrica.
10. Levantar lápis (de frente): Levantar lápis usando preensão com três dedos.
11. Levantar clipe de papel (de frente): Levantar um clipe de papel usando pinça polpa-polpa.
12. Empilhar peças de dama (de frente): Empilhar três peças de dama.
13. Virar cartas (de frente): Virar três cartas usando a pinça e supinação de antebraço.
15. Virar chave na fechadura (de frente): Utilizando a pinça da chave, virá-la para ambos os lados e voltar ao meio.
16. Dobrar toalha (de frente): Dobrar toalha longitudinalmente, em seguida, usa a mão testada para dobrar a toalha ao meio novamente.
17. Levantar a cesta (de pé): Pegar a cesta pela alça e colocá-la na superfície ao lado.