Translation, adaptation and inter-rater reliability of the administration manual for the Fugl-Meyer assessment

Tradução, adaptação e confiabilidade interexaminadores do manual de administração da escala de Fugl-Meyer*

Stella M. Michaelsen[1], André S. Rocha[1], Rodrigo J. Knabben[1], Luciano P. Rodrigues[2], Claudia G. C. Fernandes[2]

Abstract

Background: Recently, the reliability of the Brazilian version of the Fugl-Meyer Assessment (FMA) was assessed through the scoring given according to observations made by a single evaluator who applied the test. When different raters apply the scale, the reliability may depend on the interpretation given to the assessment sheet. In such cases, a clear administration manual is essential for ensuring homogeneity of application. **Objectives:** To translate and adapt the French Canadian version of the FMA administration manual into Brazilian Portuguese and to evaluate the inter-rater reliability when different evaluators apply the FMA on the basis of the information contained in the manual. **Methods:** Eighteen adults (59±10 years) with chronic hemiparesis (38±35 months after a stroke) took part in this study. Eight patients participated in the first part of the study and 10 in the second part. Based on analyzing the results from part 1, an adapted version was developed, in which information and photos were added to illustrate the positions of the patient and evaluator. The inter-rater reliability was assessed using the intraclass correlation coefficient (ICC). **Results:** The reliability of the FMA based on the adapted version of the manual was excellent for the total motor scores for the upper limbs (ICC=0.98) and lower limbs (ICC=0.90), as well as for movement sense (ICC=0.98) and upper and lower-limb passive range of motion (ICC=0.84 and 0.90, respectively). The reliability was moderate for tactile sensitivity (0.75). The joint pain assessment presented low reliability. **Conclusions:** The results showed that, except for pain assessment, application of the FMA based on the adapted version of the application manual for Brazilian Portuguese presented adequate inter-rater reliability.

Key words: translation/adaptation; motor assessment; hemiplegia/hemiparesis; reliability.

Resumo

Contextualização: Recentemente a confiabilidade da versão brasileira da Escala de Fugl-Meyer (EFM) foi avaliada pela pontuação dada pela observação de um único examinador que aplicou a escala. Quando diferentes examinadores aplicam a escala, a confiabilidade pode depender da interpretação dada à ficha de avaliação. Nesse caso, um manual de administração claro é fundamental para garantir homogeneidade na aplicação. **Objetivos:** Traduzir e adaptar para o português-Brasil a versão do Manual de Administração em francês-canadense da EFM e avaliar a confiabilidade interexaminadores quando diferentes examinadores aplicam a EFM com base nas informações contidas no manual. **Métodos:** Participaram do estudo 18 adultos (59±10 anos) com hemiparesia crônica (38±35 meses pós-Acidente Vascular Encefálico). Oito sujeitos participaram da primeira parte do estudo e dez, da segunda parte. Baseada na análise dos resultados da parte 1, desenvolveu-se uma versão adaptada à qual foram adicionadas informações e fotos para ilustrar a posição do paciente e do examinador. A confiabilidade interexaminadores foi avaliada com o Coeficiente de Correlação Intraclasse (CCI). **Resultados:** A confiabilidade da EFM baseada na versão adaptada do manual foi excelente para o escore motor total do membro superior (MS, CCI=0,98) e membro inferior (MI, CCI=0,90), sentido de movimento (CCI=0,98), amplitude de movimento (ADM) passiva do MS (CCI=0,84) e do MI (CCI=0,90) e moderada para a sensibilidade tátil (0,75). A avaliação da dor articular apresentou baixa confiabilidade. **Conclusão:** Os resultados mostram que, com exceção da avaliação da dor, a aplicação da EFM com base na versão adaptada do manual de aplicação em português-Brasil apresenta adequada confiabilidade interexaminadores.

Palavras-chave: tradução/adaptação; avaliação motora; hemiplegia/hemiparesia; confiabilidade.

[1] Postgraduate Program on Movement Sciences, Center for Health and Sports Sciences (CEFID), Universidade do Estado de Santa Catarina (UDESC), Florianópolis, SC, Brazil

[2] Physical Therapy Course, Universidade Luterana do Brasil (ULBRA), Canoas, RS, Brazil

Correspondence to: Stella Maris Michaelsen, PPG Ciências do Movimento Humano - CEFID/UDESC, Rua Pascoal Simone, 358, Coqueiros, CEP 88 080 350, Florianópolis, SC, Brazil, email: michaelsenstella@hotmail.com

Introduction ::::.

The choice of an appropriate outcome measurement is critical to the success of any study that aims to evaluate the efficacy of a proposed treatment[1]. Within this perspective, the Fugl-Meyer assessment scale (FMA)[2], which was developed in 1975, has been used both to describe the sensorimotor recovery of patients who suffered a stroke[3,4] and to classify them with regard to the severity of the sequelae[5,6]. This instrument is widely used in research to evaluate the effect of diverse treatment approaches, such as: practicing functional tasks versus strengthening[7]; Bobath versus repetitive training for the upper limbs[8]; training for unilateral tasks versus bilateral symmetrical tasks[9]; group exercises[10]; training controlled by computer versus electrical stimulation[11]; task-oriented training with or without restrictions on trunk compensatory movements[12]; and constraint-induced movement therapy versus conventional physical therapy[13]. In clinical practice, the FMA is also useful for planning and evaluating the results from a treatment[14], besides being described as one of the tests recommended for assessment of domain structure and body function within the International Classification of Functioning, Disability and Health (ICF)[15].

Recently, Maki et al.[16] developed a Brazilian version of this scale and evaluated its inter-rater reliability. Its results presented high intraclass correlation coefficients (ICC) for the total score (0.98) and its subscales (ICC between 0.94 and 0.99).

Inter-rater reliability is related to the stability of the data obtained by two or more evaluators from the same measurement. The term inter-rater reliability can be used when the measurement is obtained through observation of performance, since the rater does not have physical contact with the person under evaluation. On the other hand, the reliability of a measurement may depend on the evaluator's performance, i.e. the way in which the test is administered, when different evaluators repeat the assessment on the same patient at different times[17].

In the study by Maki et al.[16], the test was applied by a single evaluator, and the inter-rater reliability was obtained through comparison of the scores that two other professionals attributed to patients' performance. Both of these experimental designs are valuable for assessing inter-rater reliability, but when the tests require physical contact by a "hands on" evaluator, score variability can come from patients and from the evaluator's interpretation of the score sheet[14]. The reliability assessment can be named inter-rater if distinct evaluators administer the test to the same patient, in order to differentiate it from inter-observer reliability assessment, which depends only on the observation ability of the other evaluators.

In 1989, with the permission of the original author of the FMA, Dutil et al.[18] developed a manual for this test in Canadian French, which described the procedures and provided illustrative graphs, in order to facilitate test administration. To reduce the measurement bias, a transparent administration manual is crucial so that homogeneity of scale administration can be ensured, especially in tests that do not require previous training[19]. Within this context, the present study had the following goals: Part 1: a) to translate the FMA application manual developed by Dutil et al.[18] into Brazilian Portuguese; b) to investigate the reliability of the Brazilian Portuguese version translated from the manual in Canadian French; Part 2: a) to adapt the translated manual based on the limitations identified in the results from part 1; b) to investigate the inter-rater reliability of the FMA through administration by different examiners using the information contained in the translated and adapted version of the manual.

Methods ::::.

This study consisted of two parts. Part 1 was approved through Protocol 2005/113H from the Research Ethics Committee of Universidade Luterana do Brasil (ULBRA), Canoas, RS, Brazil. This part of the study included translation and administration of the manual, and assessment of its inter-rater reliability, and was carried out between May and December 2005. Next, adaptations were made to the manual, and after approval through Protocol 191/2007 from the Research Ethics Committee of Universidade do Estado de Santa Catarina (UDESC), Florianópolis, SC, Brazil, the test was performed using the same sample and independent evaluators. Part 2 of this study consisted of assessing the inter-rater reliability of the test, based on the translated and adapted version of the manual (July 2008 to May 2009).

Participants

Eighteen 18 patients (mean age of 59±10 years) who presented hemiparesis as a sequela of a stroke took part in the present study. All of the patients were recruited from clinical physical therapy schools at two institutions. A free and informed consent declaration was obtained from each participant. Subjects were included in this study if they were over 18 years of age, of either sex, presenting chronic hemiparesis (> six months after a stroke) and unilateral sequelae, and were capable of understanding the examiner's instructions for performing the FMA. Subjects with cerebral sequelae, presenting ataxia and other associated neurological diseases were excluded from the present study.

Eight subjects (aged 60.0±6.5 years) participated in part 1 of the study and ten subjects (aged 58.9±12.8 years) in part 2. The sample was defined based on the studies by Duncan, Propst and Nelson[3] and Sanford et al.[19], who assessed the reliability of the FMA. The sample size defined by Bonett[20] ranged from 15 to 21 subjects, when two evaluators administered a test with a planned ICC of 0.9, a precision estimated as 0.2 and an alpha of 0.05[21].

## Translation of the manual for FMA administration

The French-Canadian version of the manual and the application sheet for the FMA developed by Dutil et al.[18] were translated by a Brazilian physical therapist, with authorization from the author. The test presents four domains: 1) motor function; 2) sensitivity; 3) passive range of motion (ROM) and pain; and 4) balance; but the last domain was not included in the present study.

The motor function domain evaluates the capacity to perform movements of increasing complexity based on the stages for recovery proposed by Brunnstrom[22]. The scoring is based on a three-point ordinal scale: (0) cannot perform; (1) performed partially and (2) performed completely. Tendon-reflex activity, presence/absence of abnormal synergisms and the coordination and speed of voluntary movements are evaluated in the motor domain. The upper-limb subscale includes assessment of wrist movements and five types of handgrip (Figure 1). The total score on the motor scale is 100 points, such that the lower limbs are graduated from 0 to 34 points and the upper limbs, from 0 to 66.

Tactile sensitivity is evaluated using cotton wool administered to the anterior and posterior regions of the shoulder, arm, forearm, thumb and forefinger (upper limb = 20 points) and to the middle third of the anterior region of the tibia and the sole of the foot (lower limb = 4), totaling 24 points. The score for the tactile sensitivity test also ranges on an ordinal scale from 0 to 2, such that (0) is absence of sensitivity; (1) is hypo or hypersensitivity; and (2) is normal sensitivity, comparing the affected side with the contralateral. The sense of movement is assessed in the shoulder, elbow, wrist and thumb (upper limb = 8 points), and in the hip, knee and hallux (lower limb = 8 points), thus totalizing 16 points.

The scoring for the proprioceptive movement sense is defined thus: (0) the movement is not identified; (1) at least 75% of the answers are correct (three correct answers from four movements or six from eight depending on the joint evaluated); (2) all the answers are correct.
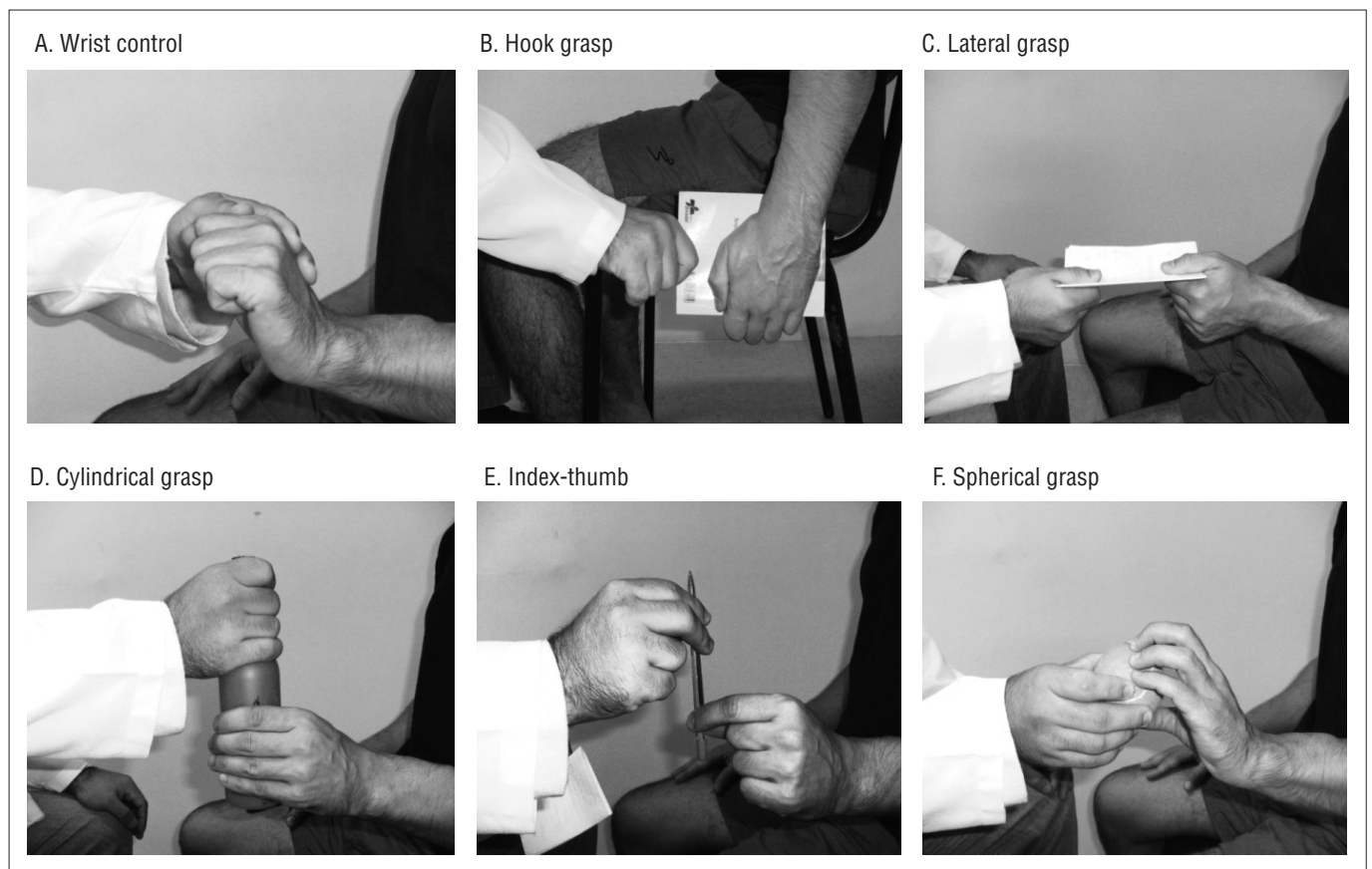

Figure 1. Items for upper limb section with examples of test to evaluate (A) wrist control, (B) hook grasp, (C) lateral grasp, (D) cylindrical grasp, (E) index-thumb grip, and (F) spherical grasp.

The range of motion (ROM) is passively evaluated in the shoulder, elbow, forearm, wrist and fingers (upper limb = 24 points) and hip, knee and ankle (lower limb = 20 points). The score for the passive ROM is defined by visual estimation in comparison with the non-paretic side, as follows: (0) some degree of ROM (less than half of the total ROM); (1) reduced ROM (maintained over more than half of the total ROM) and (2) normal ROM. The section joint pain is scored thus: (0) severe pain throughout the movement or very severe at the end of the movement; (1) little pain; (2) no pain. Both range of motion and section joint pain have a total score of 44 points[19,23].

Inter-rater reliability

In each of the parts of the study, the reliability was assessed by two different examiners, totaling four independent examiners. All the examiners were physical therapists with more than three years of experience in the neurofunctional field. The scale was administered without previous treatment and was based only on the procedures described in the administration manual. The first examiner (E1 or E3) applied the test and, after an interval of approximately one hour, the second examiner (E2 or E4) administered the test on the same patient. In the tests on motor function, the patient was first asked to perform the

A. Upper limb movement sense

B. Lower limb movement sense

C. Tactile sensibility
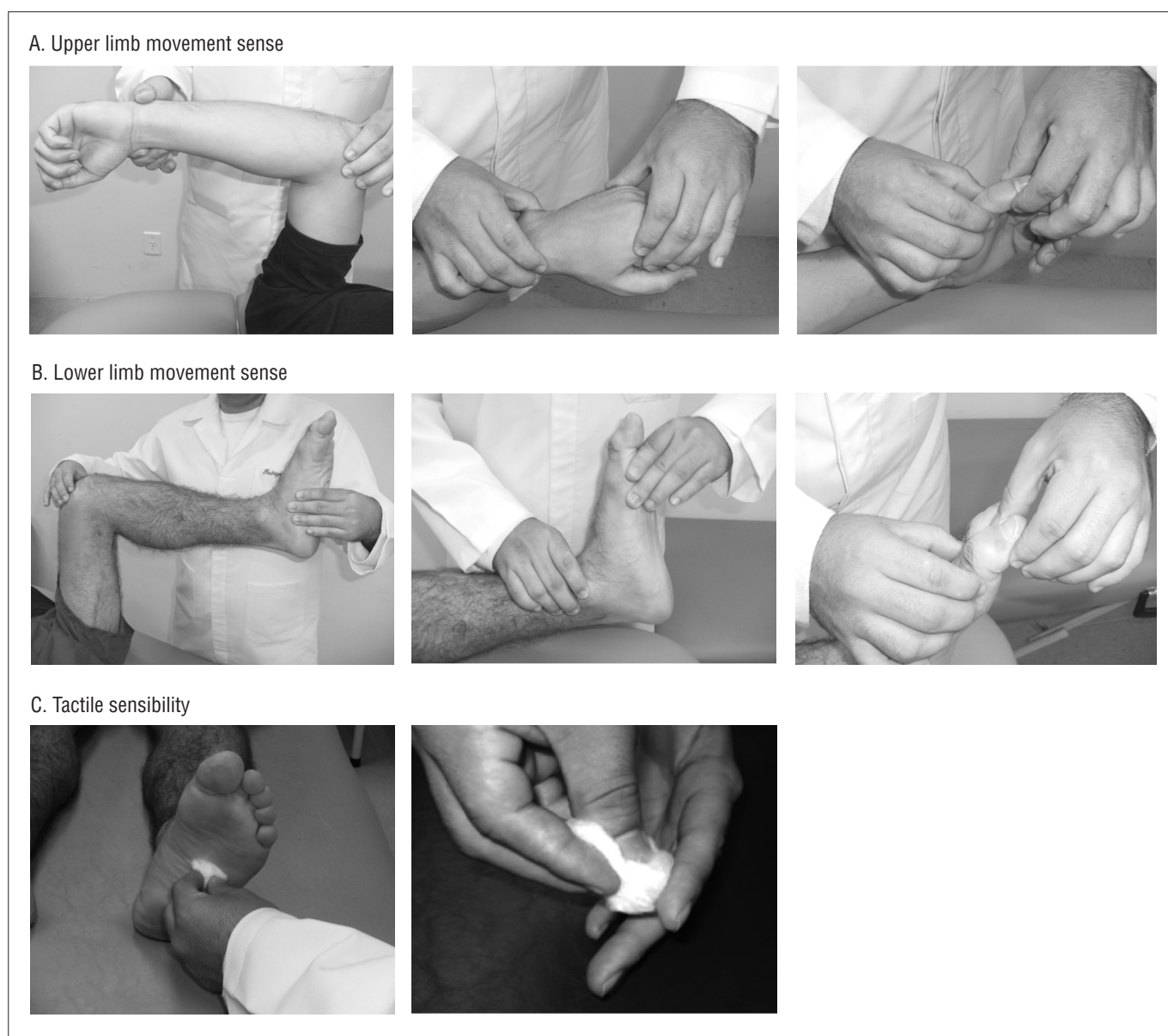
Figure 2. Items for sensory section: A) Example of test and evaluator hand placement during examination of upper limb movement sense respectively for shoulder, wrist and thumb B) Examination of lower limb movement sense respectively for hip, ankle e hallux; C) Example of cotton stimulation site during tactile sensibility examination.

tasks using the non-paretic side with the purpose of achieving comprehension, and then the test was repeated using the paretic side. The scoring was obtained in accordance with the explanations described in the manual. Thus, the procedures for assessing the reliability were similar in the two parts of the study, except that the adapted version of the manual was used in part 2.

Adaptation of the manual

From the reliability results obtained in part 1 of the study (described in the section "results"), the examiners evaluated the clarity of the manual, especially with regard to items that demonstrated inadequate or low reliability, and the suggestions were incorporated. To standardize the application of the sensitivity tests, photos illustrating the positioning and placement of the examiner's hands were incorporated in the section that evaluates movement sense, and photos of the stimulation site, in the section that evaluates tactile sensitivity (Figure 2). Photos of the positioning of the patient and examiner for administration of the reflex-tendon tests were also incorporated. In addition, the illustrations that already existed in the original manual were replaced by photos.

Statistical analysis

In order to evaluate the inter-rater reliability in parts 1 and 2 of the study, the ICC and 95% confidence interval (CI) were used. The following were compared: 1) individual items and total scores for motor function for the upper and lower limbs separately; 2) total score for tactile sensitivity; 3) total score for sense of motion; 4) passive ROM; and 5) joint pain. The following classification was used for the ICC values: weak concordance, ICC<0.40; moderate concordance, ICC≤0.75; and excellent concordance, ICC>0.75.[24]

Table 1. Demographic data on the participants and mean Fugl-Meyer Scale scores for two evaluators in each part of the study.

Characteristics	Part 1 (n = 8)	Part 2 (n = 10)
Age (years)	60.0 (± 6.5)	58.9 (± 12.8)
Time elapsed since stroke (months)	39.0 (± 33.6)	36.3 (± 37.7)
Sex (M/F)	7/1	5/5
Fugl-Meyer Scale		
Motor Function (100)	65.1 (± 22.9)	62.7 (± 26.7)
Tactile Sensitivity (24)	18.8 (± 4.8)	18.4 (± 7.7)
Movement sense (16)	12.6 (± 2.9)	14.3 (± 3.0)
Passive ROM (44)	37.1 (± 1.9)	38.8 (± 4.6)
Joint Pain (44)	36.7 (± 5.5)	37.7 (± 6.6)

M = male; F = female; ROM = range of motion.

Results

The demographic characteristics and level of impairment of the participants (obtained as the mean score given by the two examiners) were similar in the two parts of the study and are described in Table 1.

The level of motor impairment evaluated by the motor function of the FMA ranged from severe (21 points) to slight (96 points), considering the two samples studied. In the second part of the study, four participants had severe impairment (≤50 points); three, significant impairment (51-84 points); and two, moderate impairment (≥96 points) in the FMA[2].

Inter-rater reliability – Part 1

In the first part of the study, the mean values (±SD) obtained, respectively, by E1 and E2 for the motor scores were: total (63.0±23.5 and 67.2±23.7); upper limbs (36.5±21.4 and 38.9±20.3) and lower limbs (26.5±4.2 and 28.4±4.6). The ICCs between the two raters for the total score and upper limb score were 0.99 (p<0.001) for both and 0.88 (p=0.006) for the lower limbs. Except for reflex activity (ICC=0.71), the inter-rater reproducibility was excellent for all the other sub-items of the motor section that evaluated the upper limbs (ICC between 0.90 and 0.99). For the lower limbs, the items of synergy motion (0.84) and the coordination/speed test (0.90) showed excellent reliability. Flexor/extensor synergy and movements combining the synergy reached, respectively, moderate reliability (0.68) and low reliability (0.38). The subitem of reflexes obtained the lowest ICC (0.27). The ICC between the two examiners for tactile sensitivity, movement sense, passive ROM of the lower limbs and pain showed reliability ranging from low to moderate. Only the passive ROM for the upper limbs presented excellent reliability (Table 2).

Inter-rater reliability – Part 2

The inter-rater reliability for the sensitivity assessment (tactile sensitivity and sense of motion), ROM and joint pain values from parts 1 and 2 of the study are shown in Table 2. Comparison between the ICCs from parts 1 and 2 for the items of tactile sensitivity, movement sense and ROM shows that the reliability became excellent after the adaptation of the manual. The reliability of the joint pain assessment for the upper limbs was moderate, and for the lower limbs was low (Table 2).

In part 2 of the study, the ICC between two examiners for the total motor function score was 0.98 (p<0.001). The mean values (±SD) obtained by E3 and E4, and the ICC and 95% CI

Table 2. Descriptive statistics for sensitivity, passive range of motion (ROM) and joint pain items in Parts 1 and 2 of the study. Comparison between inter-rater reliability of the French-Canadian Fugl-Meyer Assessment manual translated into Brazilian Portuguese (Part 1) and manual with adaptation (Part 2).

Study	Part 1			Part 2		
	R1 Mean (SD)	R2 Mean (SD)	ICC (95% CI)	R3 Mean (SD)	R4 Mean (SD)	ICC (95% CI)
Tactile Sensitivity for UL and LL (24)	18.4 (4.7)	19.1 (5.3)	0.58	18.5 (7.2)	18.2 (8.5)	0.75* (0.02-0.94)
Movement Sense for UL and LL (16)	10.3 (2.1)	15.0 (1.2)	-0.87	14.4 (3.1)	14.1 (3.1)	0.98*** (0.90-0.99)
Passive ROM for UL (24)	20.3 (2.3)	18.4 (1.8)	0.76* (0.19-0.95)	21.7 (2.2)	21.1 (2.6)	0.84** (0.37-0.96)
Passive ROM for LL (20)	17.4 (1.8)	18.3 (0.7)	0.50	17.5 (3.2)	17.2 (3.5)	0.90*** (0.60-0.97)
Joint pain for UL (24)	16.4 (3.1)	20.4 (2.7)	0.59	22.7 (1.9)	17.3 (5.6)	0.58
Joint pain for LL (20)	16.6 (4.0)	20.0 (0.0)	0.00	19.2 (1.9)	16.2 (2.7)	0.06

R = rater; SD = standard deviation; ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; UL = upper limb; LL = lower limb. *p < 0.05; ** p < 0.01; ***p < 0.001.

Table 3. Descriptive statistics, inter-rater reliability and 95% confidence intervals for individual items for both upper and lower-limb motor function after manual adaptation (final version).

	Rater 3 Mean (SD)	Rater 4 Mean (SD)	Inter-rater ICC	(95% CI)
Motor Function individual items for UL				
Reflex activity §	4.2 (1.1)	4.7 (0.8)	-0.74	(0.02-0.57)
Flexor and extensor synergies	9.0 (6.9)	10.8 (6.4)	0.93***	(0.73-0.98)
Motion with mixed synergy	3.5 (2.7)	3.9 (2.3)	0.97***	(0.88-0.99)
Motion without synergy	2.2 (2.3)	2.6 (2.5)	0.96***	(0.83-0.99)
Wrist control	4.0 (4.1)	4.8 (4.3)	0.98***	(0.94-0.99)
Manual control	8.9 (5.5)	9.0 (5.5)	0.96 ***	(0.82-0.99)
Coordination/speed	3.0 (2.7)	2.9 (2.7)	0.98***	(0.92-0.99)
Total UL	34.8 (22.6)	38.8 (22.9)	0.98 ***	(0.94-0.99)
Motor Function individual items for LL				
Reflex activity §	4.8 (0.9)	4.5 (0.8)	0.24	(2.02 -0.81)
Flexor and extensor synergies	10.9 (2.8)	11.6 (2.5)	0.60*	(0.01-0.88)
Motion with mixed synergy	2.9 (1.3)	3.5 (1.0)	0.84**	(0.36-0.96)
Motion without synergy	2.4 (1.4)	2.9 (1.3)	0.85**	(0.41-0.96)
Coordination/speed	4.5 (1.8)	4.8 (1.8)	0.91***	(0.63-0.97)
Total LL	24.8 (6.2)	26.9 (6.1)	0.90 ***	(0.61-0.97)

ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; §reflex activity including sum of items 1 and 5; UL = upper limb; LL = lower limb; *p < 0.05; ** p < 0.01; ***p < 0.001.

for the individual items of the motor function assessment, are presented in Table 3. The item of reflex activity presented low reliability both for the upper and for the lower limbs. The ICCs for the individual items in the assessment of voluntary motor function of the upper limbs ranged from 0.93 to 0.98, i.e. reaching excellent concordance. The same excellent result was found for the subitems of lower-limb motor function, except for flexor synergy, which obtained a moderate value (0.60). Excellent reliability was found for the total scores for upper and lower-limb motor function.

Discussion

The FMA has been widely used in both research and clinical practice[10,12,14], and is suitable for evaluating domain structure and body function within the context of the ICF[15]. Use of a reliable scale, applied on the basis of information contained only in the manual, is advantageous because this makes it possible to minimize the training needs. This characteristic, in addition to allowing easy access, confers greater autonomy for clinicians in assessing the motor recovery of subjects with hemiparesis.

Considering the high prevalence of hemiparesis, it is essential that physical therapists are aware of the proper methods for administering and scoring scales that assess recovery after a stroke. The results from this study demonstrated that this version in Brazilian Portuguese, adapted from the manual developed by Dutil et al.[18], presents sufficient information to enable use of the FMA by different examiners with adequate reliability.

The inter-rater reliability of the total values for the FMA and for the upper and lower limb sections were excellent and in agreement with the test in its original version[3,25]. Among a sample of subacute patients (up to six months post-stroke), Sanford et al.[19] evaluated the inter-rater reliability between three physical therapists who applied the test at one-day intervals. The ICC for the total score was 0.96 and for the upper and lower-limb subscales, it was 0.97 and 0.92, respectively. The adaptation of the manual provided an increase in the reliability of the sensory evaluation and the passive ROM evaluation, but the subitems of reflex activity and joint pain showed low reliability, even after implementing the final version. The source of test variability may differ according to the domain evaluated[19]. From this perspective, in order to assist readers in judging the adequacy of this scale for its use in clinical practice and research, and to assess the particular aspects of recovery after a stroke, the reliability results will be discussed separately for the subitems.

Using the original version of the FMA[2], and with previous training for the examiners, Sanford et al.[19] found high values (ICC=0.85) for ROM and sensitivity and moderate reliability (ICC=0.66) for the joint pain domain. Our results reached only moderate levels for the upper limbs, while the reliability of the pain assessment in the lower limbs was low. Sanford et al.[19] indicated that within the field of joint pain, examiners assign scores in accordance with the examinee's ability to describe his/her feelings, and that it may be difficult to assess the presence of aphasia in these domains. The instructions in the manual did not seem to contribute towards achieving adequate reliability in pain assessments made by different examiners. However, the reliability of the pain assessment may be influenced by the inherent variability of the measurement. Another important factor is that the domains of pain and sensitivity are highly dependent on the patient's attention and cooperation[26]. Therefore, one limitation to application of this scale would be in relation to using it for patients with comprehension aphasia, attention deficits and cognitive impairments.

Lin et al.[27] exclusively assessed the psychometric properties of the sensory scale of the FMA, as implemented by two examiners at a maximum interval of 48 hours, among patients in the acute post-stroke phase. In a sample of 176 patients, the reliability of the sensitivity domain (including light touch and proprioception) was 0.93. The agreement between the examiners for the eight items that assessed proprioception was good to excellent (weighted kappa between 0.71 and 0.90). In the same study, the use of only four locations (maximum 8 points) for assessment of tactile sensitivity resulted in agreement ranging from poor (0.30) to moderate (0.55). The low agreement for tactile sensitivity found in the study by Lin et al.[27] may have been influenced by the participants' level of attention. Among patients presenting acute changes in attention, the applicability of the sensory scale is limited. The small number of sites tested by Lin et al.[27] may also have influenced the results. The administration manual developed by Dutil et al.[18] describes the evaluation of ten sites in the upper limbs (five regions on the ventral and five on the dorsal face of the paretic upper limb) and two for the paretic lower limb (anterior leg and plantar surface of the foot), with a maximum score of 24 points. Malouin et al.[28] compared the FMA with the motor scale of the Motor Assessment Scale, using the same manual as in this study, but the reliability values for the FMA in this domain were not presented. Our findings suggest that evaluation of tactile sensitivity at a larger number of sites and the use of photos to standardize the application sites seem to increase the reliability of the tactile subscale.

The section of the FMA most cited in the literature has been the upper-limb motor function. When the score has been calculated from video recordings, the total scores for the upper limbs in the FMA have shown excellent reliability, with ICC of 0.99[16,21]. With two different examiners applying the test, the reliability of the total score for the upper limbs in the FMA based on the Brazilian version of the manual remained excellent (ICC=0.98), with values similar to the findings of Sanford et al.[19], in which the reliability of the English version was assessed by three different examiners. Regarding the subitems, our results for the upper limbs showed that only reflex activity had poor reproducibility. Similarly, Duncan, Propst and Nelson[3], in a study on 18 chronic patients, found significant differences between four raters for the item of reflex activity, using repeated-measurement ANOVA to compare the scores. Recently, in a study on the dimensionality and validity of the FMA for the upper limbs, Woodbury et al.[29] suggested that the item of reflex activity seemed not to contribute towards the construct of motor recovery. The results from Stam and Van Crevel[30] regarding the reliability of tendon reflexes showed considerable discrepancy between their three raters. In addition, the score for the reflex scale differed from the other scores. While the motor function, sensitivity and ROM were scored according to a three-point ordinal scale, the reflexes in item 1 were scored on a two-point scale, on which 0 = absent and 2 = present[19]. Platz et al.[21] evaluated the reliability of the FMA in relation to the upper-limb section alone, from the video analysis, and

found ICC=0.99 for the total score. They did not present the scores for the reliability evaluation on reflex activity separately, thus confirming that when inserted in the maximum score, the item of reflex activity does not appear to decrease the reliability of the total score for this subscale.

The total score for lower-limb motor function in the FMA showed excellent reliability (ICC=0.90) when different examiners applied the scale. In the study by Sanford et al.[19], the ICC was 0.92, thus confirming that the ICC for the lower limbs in the FMA is generally slightly lower than the ICC for the upper limbs, but that the reliability is still excellent. The analysis on the lower-limb subitems showed that mixed movements with and without synergy, as well as coordination/speed measurements, had excellent reproducibility, which was not observed for the subitems of flexor and extensor synergy, which showed moderate reliability. The study by Duncan, Propst and Nelson[3] also showed that there were no significant differences in the scores given by different examiners for the lower limbs. Just as for the upper limbs, the items that assessed reflex activity did not show adequate reliability. Within this context, evaluations on the tendon reflex of the quadriceps and biceps muscles may have coefficients of variation of respectively 54% and 60% for different evaluators[31,32]. Nonetheless, when the findings for the upper limbs are added to the total score, the low reliability of reflex measurements does not seem to interfere with the reliability of the lower-limb subscale.

Although the results from different examiners' applications of the FMA based on the manual were shown to be reliable, one of the limitations of this study was the small sample size. However, the only items that did not show moderate or high ICC in Part 2, after the adaptation of the manual, were pain and reflexes. Low levels of reliability were found for these same items in the FMA in studies with larger samples[18,27]. Thus, it would be unlikely that this result was due to a type II error.

The results from the present study showed that the motor subscales for the upper and lower limbs, sense of movement and passive ROM of the Brazilian version of the FMA, when administered on the basis of information contained in the administration manual, had excellent inter-rater reliability. Tactile sensitivity presented moderate reliability. Although the items of reflex activity and joint pain did not achieve adequate reliability, the manual is generally sufficient for guiding correct use of the FMA in clinical practice. It can be used to evaluate the recovery of patients after a stroke.

References ::::.

1. Good DC. Stroke: promising neurorehabilitation interventions and steps toward testing them. Am J Phys Med Rehabil. 2003;82(10 Suppl):S50-7.

2. Fugl-Meyer AR, Jääskö 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.

3. Duncan PW, Propst M, Nelson SG. Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. Phys Ther. 1983;63(10):1606-10.

4. Hsieh YW, Hsueh IP, Chou YT, Sheu CF, Hsieh CL, Kwakkel G. Development and validation of a short form of the Fugl-Meyer motor scale in patients with stroke. Stroke. 2007;38(11):3052-4.

5. Plummer P, Behrman AL, Duncan PW, Spigel P, Saracino D, Martin J, et al. Effects of stroke severity and training duration on locomotor recovery after stroke: a pilot study. Neurorehabil Neural Repair. 2007;21(2):137-51.

6. Michaelsen SM, Natalio MA, Silva AG, Pagnussat AS. Confiabilidade da tradução e adaptação do 'Test d'Évaluation des Membres Supérieurs des Personnes Agées' (TEMPA) para o português e validação para adultos com hemiparesia. Rev Bras Fisioter. 2008;12(6):511-9.

7. Winstein CJ, Rose DK, Tan SM, Lewthwaite R, Chui HC, Azen SP. A randomized controlled comparison of upper-extremity rehabilitation strategies in acute stroke: A pilot study of immediate and long-term outcomes. Arch Phys Med Rehabil. 2004;85(4):620-8.

8. Platz T, Eickhof C, van Kaick S, Engel U, Pinkowski C, Kalok S, et al. Impairment-oriented training or Bobath therapy for severe arm paresis after stroke: a single-blind, multicentre randomized controlled trial. Clin Rehabil. 2005;19(7):714-24.

9. Desrosiers J, Bourbonnais D, Corriveau H, Gosselin S, Bravo G. Effectiveness of unilateral and symmetrical bilateral task training for arm during the subacute phase after stroke: a randomized controlled trial. Clin Rehabil. 2005;19(6):581-93.

10. Pang MY, Harris JE, Eng JJ. A community-based upper-extremity group exercise program improves motor function and performance of functional activities in chronic stroke: a randomized controlled trial. Arch Phys Med Rehabil. 2006;87(1):1-9.

11. Hesse S, Werner C, Pohl M, Rueckriem S, Mehrholz J, Lingnau ML. Computerized arm training improves the motor control of the severely affected arm after stroke: a single-blinded randomized trial in two centers. Stroke. 2005;36(9):1960-6.

12. 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.

13. Wu CY, Chen CL, Tang SF, Lin KC, Huang YY. Kinematic and clinical analyses of upper-extremity movements after constraint-induced movement therapy in patients with stroke: a randomized controlled trial. Arch Phys Med Rehabil. 2007;88(8):964-70.

14. Finch E, Brooks D, Stratford PW, Mayo NE. Physical Rehabilitation Outcome Measures - A Guide to Enhanced Clinical Decision Making. 2nd. Canadian Physiotherapy Association; 2002.

15. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J. Issues for selection of outcome measures in stroke rehabilitation: ICF Body Functions. Disabil Rehabil. 2005;27(4):191-207.

16. Maki T, Quagliato EMAB, Cacho EWA, Paz LPS, Nascimento NH, Inoue MMEA, et al. Estudo da confiabilidade da aplicação da escala de Fugl-Meyer no Brasil. Rev Bras Fisioter. 2006;10(2):117-83.

17. Marx RG, Bombardier C, Wright JG. What do we know about the reliability and validity of physical examination tests used to examine the upper extremity? J Hand Surg Am. 1999;24(1):185-93.

18. Dutil E, Arsenault AB, Corriveau H, Prévost R. Protocole d'évaluation de la fonction sensori-motrice: Test de Fugl-Meyer. Montreal, Canada: La librairie de l'Université de Montréal; 1989.

19. 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.

20. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. Stat Med. 2002;21(9):1331-5.

21. Platz T, Pinkowski C, van Wijck F, Kim IH, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. Clin Rehabil. 2005;19(4):404-11.

22. Brunnstrom S. Motor testing procedures in hemiplegia: based on sequential recovery stages. Phys Ther. 1966;46(4):357-75.

23. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehabil Neural Repair. 2002;6(3):232-40.

24. Fleiss JL. Statistical Methods for rates and proportions. 2nd ed. New York: John Wiley & Sons; 1999.

25. Berglund K, Fugl-Meyer AR. Upper extremity function in hemiplegia. A cross validation study of two assessment methods. Scand J Rehabil Med. 1986;18:155-7.

26. Harris RE, Williams DA, McLean SA, Sen A, Hufford M, Gendreau RM, et al. Characterization and consequences of pain variability in individuals with fibromyalgia. Arthritis Rheum. 2005;52(11):3670-4.

27. Lin JH, Hsueh IP, Sheu CF, Hsieh CL. Psychometric properties of the sensory scale of the Fugl-Meyer Assessment in stroke patients. Clin Rehabil. 2004;18(4):391-7.

28. Malouin F, Pichard L, Bonneau C, Durand A, Corriveau D. Evaluating motor recovery early after stroke: comparison of the Fugl-Meyer Assessment and the Motor Assessment Scale. Arch Phys Med Rehabil. 1994;75(11):1206-12.

29. Woodbury ML, Velozo CA, Richards LG, Duncan PW, Studenski S, Lai SM. Dimensionality and construct validity of the Fugl-Meyer Assessment of the upper extremity. Arch Phys Med Rehabil. 2007;88(6):715-23.

30. Stam J, van Crevel H. Measurement of tendon reflexes by surface electromyography in normal subjects. J Neurol. 1989;236(4):231-7.

31. Toft E, Sinkjaer T, Rasmussen A. Stretch reflex variation in the relaxed and the pre-activated quadriceps muscle of normal humans. Acta Neurol Scand. 1991;84(4):311-5.

32. Segal RL, Wolf SL. Variability of human biceps brachii spinal stretch reflexes: control conditions. J Electromyogr Kinesiol. 1993;3(1):24-32.