ORIGINAL REPORT

CLINICAL TESTING OF AN INNOVATIVE TOOL FOR THE ASSESSMENT OF BIOMECHANICAL STRATEGIES: THE TIMED "UP AND GO" ASSESSMENT OF BIOMECHANICAL STRATEGIES (TUG-ABS) FOR INDIVIDUALS WITH STROKE

Christina D. C. M. Faria, PT, PhD[1, 2], Luci F. Teixeira-Salmela, PT, PhD[1] and Sylvie Nadeau PT, PhD[2]

From the [1]Department of Physical Therapy, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil and [2]Centre de Recherche Interdisciplinaire en Réadaptation (CRIR), Institut de Réadaptation Gingras-Lindsay-de-Montréal, Université de Montréal, Faculté de Médecine, École de réadaptation, Montréal, Québec, Canada

Objective: To investigate the reliability and construct and criterion-related validities of the Timed "Up and Go" Assessment of Biomechanical Strategies (TUG-ABS), when used with subjects with hemiparesis due to stroke within clinical settings.
Design/methods: Construct validity was investigated by the following methods: the known groups, convergence, discriminant analyses, and the opinions of clinical professionals, who used the TUG-ABS with subjects with stroke. The criterion-related validity was investigated by comparing the real-time and video observation scores. Inter-rater reliability was investigated by two independent examiners using both real-time and video observations.
Results: The TUG-ABS differentiated people with stroke from healthy controls ($p < 0.001$), was correlated with the time spent to perform the TUG ($r_s = -0.85$; $p < 0.001$), and correctly classified 98% of the subjects with stroke ($p < 0.001$). In addition, all of the clinicians who used the TUG-ABS in their clinical settings, provided positive evaluations. Agreement was also observed between real-time and video observations ($0.27 \leq$ kappa ≤ 0.85; $p < 0.01$). Furthermore, the TUG-ABS was reliable for both real-time ($0.24 \leq$ kappa ≤ 1.00; $p < 0.05$) and video observations ($0.15 \leq$ kappa ≤ 0.94; $p < 0.05$).
Conclusion: The TUG-ABS demonstrated good construct and criterion-related validities, as well as reliability, when applied in subjects with stroke within clinical settings, which supported the theoretical assumptions employed for its development.

Key words: stroke; assessment; construct validity; clinimetric properties; mobility.

INTRODUCTION

Validity is related to the extent to which an instrument measures what it is intended to measure (1, 2). Validation of an instrument is a continuing process and, therefore, is not a property of the test or assessment (3). In addition, validity is not inherent to an instrument and should be investigated within the context of the test's intended use and specific populations (1).

Traditionally, 4 kinds of validity are described: face, content, criterion-related, and construct (1, 2). All of these types are related to the meaning of the test scores and to the inferences, which could be elaborated based upon the provided measurements (3). However, among the 4 types, construct validity is considered to be the most robust, since it allows for the establishment of the degree to which the instrument reflects the theoretical components of the construct that it intends to measure (1, 2). As stated by Messick (3), the principles of construct validity apply to all assessments, including performance assessments. However, for observational performance-based tests already developed for people with stroke, there is a lack of information regarding their construct validity (4–8).

Recently, a clinically-oriented tool, the Timed "Up and Go" Assessment of Biomechanical Strategies (TUG-ABS) (9), was developed to identify the biomechanical strategies adopted by people with stroke during the performance of the Timed "Up and Go" (TUG) test. The TUG-ABS was developed to more comprehensively evaluate functional mobility of subjects with stroke, by systematically evaluating changes in biomechanical strategies during the performance of the TUG's sequential activities in clinical practice (9). It was anticipated that this information could enhance clinical treatment planning.

The development and validation of the TUG-ABS has been recently described (9) and its reliability has been established prior to further investigation of other psychometric properties. The investigation of reliability in the initial process of instrument development is a robust process to determine which items should be retained, revised, or excluded. Further psychometric testing can then be conducted on a reliable preliminary version (1, 10, 11). However, it is necessary to investigate the reliability of the measurements of the TUG-ABS within the context of the test's intended use.

A systematic, clear, and objective process of investigation of the concurrent criterion-related validity of the TUG-ABS applied to subjects with stroke was carried out and resulted in

the establishment of its final version (9). In a previous study (9), criterion-related validity was established by video analyses to minimize biases (1). However, since the TUG-ABS was designed to be used within clinical contexts, it is necessary to determine if real-time measurements would also be valid.

Therefore, the aim of the present study was to investigate the construct and criterion-related validities, as well as the reliability of the TUG-ABS, when used within clinical settings.

METHODS

The construct validity of the TUG-ABS was investigated by both traditional (Study 1) and contemporary methods (Study 2). The traditional methods included the investigation of known groups, convergence, and discriminant analyses (1). The contemporaneous method followed previous definitions and instructions provided by Messick (3) and involved the analyses of the opinions of clinical professionals, who used the TUG-ABS with their patients with stroke in their clinical settings. The criterion-related validity was investigated by comparing the real-time and video observation scores (Study 3). The inter-rater reliability was investigated by two independent examiners during both real-time and video observations (Study 3).

Study 1: Construct validity by traditional methods

Individuals with hemiparesis due to stroke were recruited, following the characteristics of the target population for the use of the TUG-ABS: individuals from the general community with motor impairments, characterized by residual weaknesses and/or increased tonus of the paretic lower limb; who had the ability to follow instructions; were able to perform the TUG with or without assistive devices; and were over 20 years of age. People with receptive aphasia were excluded. To determine the presence of residual weaknesses (strength differences between the lower limbs greater than 15%) and/or increased tonus (modified Ashworth scale scores greater than zero) of the paretic knee extensor muscles, a handheld dynamometer (Microfet 2®, Hoggan Health Industries Draper, Utah, USA) and the modified Ashworth scale were respectively used, following established protocols (12, 13).

Individuals with stroke were divided into 3 sub-groups based on their TUG performance levels (fast, moderate, and slow), which were determined considering the reference values of the times, by calculating the 95% confidence intervals (CIs) of the mean time to complete the TUG, previously shown to differentiate between patients with stroke with mild, moderate, and severe neurological impairments (14).

Healthy control individuals, matched by age, gender, and levels of physical activity, with no histories of health problems that could affect their TUG performances were also recruited. As recommended by the Physical Activity Trends/United States (15), the levels of physical activity (inactive, insufficient, moderate, and vigorous) were determined according to the frequency, duration, and intensity of the estimated metabolic expenditures of the exercise usually performed by the individuals.

Prior to data collection, eligible participants were informed of the objectives of the study and provided consent, which was approved by the University Research Ethical Review Board. Demographic and clinical data were collected by the same physical therapist (PT). The subjects then sat in a chair (depth 45 cm, width 49 cm, arm rest height 20 cm) (16), whose height was adjusted to 100% of their leg length and the back rest adjusted to the trunk position of 90º (7, 17), to perform the TUG. They were instructed to sit comfortably with their backs against the chair, and on the word "go", stood up, walked at a self-selected comfortable speed over the 3-metre mark, turned around, walked back and sat down in the chair, as usually performed in the TUG test (18). After a familiarization trial, the TUG was performed. If there was a risk of falling, the examiner followed half a step behind the subjects, so as not to influence their walking pace (16, 19).

Three video cameras (Sony TRV 950®, Sony HC40® and Sony DCR-DVD408) were used to record the subjects' TUG performances. They were synchronized and positioned in the frontal plane, and left and right sagittal planes. Only 1 TUG performance was recorded for each subject and the video was processed and edited with Adobe® After Effects CS3® software, which allowed the 3 views to be grouped into the same file (20). This meant that all 3 views could be observed simultaneously on one screen. To avoid biases related to memory, the subjects' face was pixelated through Adobe® After Effects CS3® (20) and the videos were shown randomly for each observation session (1, 10, 11, 21, 22).

An independent examiner, after a period of familiarization with the TUG-ABS, randomly observed the videos of individuals with and without stroke. Videos were observed at normal speeds, without stopping or slowing movements, with as many trials as necessary to score all of the items.

Statistical analyses

Descriptive statistics were used for characterization purposes. χ^2, Mann-Whitney U and independent Student t-tests were employed to verify whether the healthy and stroke groups were correctly matched by age, gender, and levels of physical activity.

For the known groups' analyses, the Mann-Whitney U test was used to compare the TUG-ABS scores between the healthy and stroke groups, whereas independent Student t-tests were employed to compare the TUG times. One-way analysis of variance (ANOVAs), followed by Bonferroni *post-hoc* tests, were employed to assess whether the TUG times were different between the 3 stroke sub-groups.

For the convergence analyses, Spearman's correlation coefficients were calculated to determine the associations between the TUG-ABS scores and the time spent by people with stroke to perform the TUG.

For the discriminant analyses, two models were used to investigate whether the TUG-ABS scores could predict group membership between the stroke and control groups, as well as stroke sub-group membership (slow, moderate, and fast TUG performances) (23). With the regression analyses, discriminant functions can be used for descriptive and predictive purposes (1, 24, 25) and, in the present study, the latter was the point of interest.

All of the statistical analyses were performed with SPSS® for Windows (version 13.0), with a significance level of 5%.

Study 2: Construct validity by contemporary methods

As stated by Messick (3), "… what needs to be valid is the meaning of the interpretation of the score" and "… score validation is an empirical evaluation of the meaning and consequences of the measurement". Furthermore, "… in its simplest terms, construct validity is the evidential basis for score interpretation". Therefore, following previous definitions and instructions provided by Messick (3), construct validity by contemporary methods was also investigated for the TUG-ABS.

Clinical PTs from the city of Belo Horizonte, Brazil, involved in rehabilitation of people with stroke and who, therefore, could be assigned as the target group to use the TUG-ABS (3, 10, 11), were invited to participate. They also provided consent, which was approved by the University Research Ethical Review Board, and received explanations regarding all of the processes involved in the development of the TUG-ABS. In addition, they were instructed to familiarize themselves with the TUG-ABS.

All of the PTs who agreed to participate used the TUG-ABS with their patients with stroke who had the same characteristics as the previously described target population for the TUG-ABS. They were then asked to reply to a semi-structured questionnaire, which sought their opinions regarding different aspects of construct validity, as suggested by Messick (3), that are also related to the clinimetric properties of the TUG-ABS (Appendix I).

Statistical analyses

From questions A to D of the semi-structured questionnaire, response categories equal to 3 or 4 were considered as showing adequate con-

struct validity evaluation (26). Therefore, the frequencies of responses equal to 3 or 4 were obtained and the cumulative percentages were calculated. For question E, the frequency of response category equal to 3 was obtained and the cumulative percentage was also calculated (26).

Study 3: Criterion-related validity and reliability

Individuals with stroke were recruited from the general community, following the characteristics of the previously described target population for use of the TUG-ABS. Eligible participants were informed of the objectives of the study and were asked to provide consent, which was approved by the University Research Ethical Review Board. Demographic and clinical data were collected by the same PT.

The subjects performed the TUG, as previously described, while two independent examiners, after a period of familiarization with the TUG-ABS, observed their performance directly. The examiners selected the position that they judged to be the most effective for observation. They were instructed to score each TUG-ABS item independently. Subjects performed the TUG as many times as necessary for all of the items to be scored. All of the procedures regarding the TUG performances (explanation, familiarization, "go command") were performed by a third examiner, who also instructed the subjects to try to perform each TUG trials in the same way.

Three video cameras recorded the subjects' performances. Four weeks later, the same examiners, who had previously scored the TUG-ABS items, observed the recorded videos of the same trials in a random order and, once again, scored the TUG-ABS items independently. Stopping or slowing the videos was not allowed. The videos were repeated until all items were scored.

Statistical analyses

Descriptive statistics were used for characterization purposes. For criterion-related validity, unweighted *kappa* statistics were used to verify the levels of absolute agreement for each item score obtained by real-time and videos observations by each independent examiner. This statistical test addressed the extent to which the raters essentially reproduced the same scores and considered all disagreements in ratings with equal weighting (22, 27).

For the reliability analyses, between-rater agreement levels for both real-time and video observations were tested by quadratic weighted *kappa* statistics, which are more appropriate to assess levels of agreement on ordinal scales (22).

For all kappa statistics, κ values were interpreted as follows: below 0 as less than chance, 0.01–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as good and above 0.80 as very good levels of agreement (28, 29). All analyses were performed with StatsDirect®, 2.7.2 version for Windows, with a level of significance of 5%.

RESULTS

Study 1: Construct validity by traditional methods

Forty-eight individuals with stroke participated, 24 men and 24 women, with a mean age of 59.3 years (standard deviation (SD) 15.8 years) (range 23–90 years) and a mean stroke onset of 51.3 months (SD 49.9 months) (range 1–161 months). In terms of physical activity, 25 were inactive, 18 had insufficient levels, 2 had moderate levels and 3 had vigorous levels. Forty-eight healthy control subjects also participated, 24 men and 24 women, with a mean age of 59.1 years (SD 15.8 years) (range 23–86 years), 25 were inactive, 18 had insufficient levels, 2 had moderate levels and 3 had vigorous levels. The stroke and control groups were similar in terms of age, gender, and levels of physical activity ($p > 0.05$).

The mean TUG times for the healthy and stroke groups were, respectively, 9.5 s (SD 2.4 s) and 22.0 s (SD 16.1 s)

Table I. *Results of the discriminant analyses for classifying group membership according to the TUG-ABS scores (*n = 96*)*

Actual group	Healthy n (%)	Stroke n (%)
Healthy (*n*=48)	33 (68.8)	15 (31.3)
Stroke (*n*=48)	1 (2.1)	47 (97.9)

TUG-ABS: Timed "Up and Go" Assessment of Biomechanical Strategies.

($t = 5.3$; $p < 0.001$). Of the 48 individuals with stroke, 25 were classified as fast performers, with a mean TUG time of 12.6 s (SD 2.4 s); 14 as intermediate, with a mean time of 21.0 s (SD 3.9 s); and 9 as slow, with a mean time of 49.7 seconds (SD 18.3 seconds). ANOVAs revealed significant differences in TUG times between the stroke sub-groups ($F = 119.5$; degrees of freedom (df) = 2; $p < 0.001$) and Bonferroni post-hoc tests revealed significant differences between all of the 3 stroke sub-groups ($0.004 \leq p \leq 0.001$).

The mean number of video repetitions required to observe and score all of the items of the TUG-ABS by video observations was 3.7 repetitions (SD 1.0) (range 1–7) and 3.6 repetitions (SD 1.2) (1–7) on the first and second days of evaluation, respectively. The median TUG-ABS score of individuals with stroke was 16 (SD 2) and 24 (SD 8) for controls: the stroke group had significantly lower values than those of the healthy group ($p < 0.001$). Significant and negative correlations were found between the TUG-ABS scores and the time spent to perform the TUG ($r_s = -0.85$; $p < 0.001$).

A significant discriminant function of the TUG-ABS to predict group membership (people with and without stroke) with the TUG-ABS scores was found, with a canonical correlation of 0.68 (Wilks λ = 0.54, $\chi^2 = 57.31$, $p < 0.001$). The overall accuracy of the classification was 83.3%, indicating an exact classification for the majority of the originally grouped cases. As can be observed in Table I, 97.9% of the subjects with stroke were correctly classified to the originally grouped cases.

A significant discriminant function of the TUG-ABS to predict stroke sub-group membership (slow, moderate, and fast TUG performances) was also found, with a canonical correlation of 0.86 (Wilks λ = 0.19, $\chi^2 = 63.50$, $p < 0.001$). The overall accuracy of the classification was 83.3%, indicating that the prediction of group membership regarding the TUG performance was correctly classified for the majority of the originally grouped cases. As can be observed in Table II, 88% of the subjects of the fast sub-group, 78.6% of the moderate sub-group and 77.8% of the slow sub-group were correctly classified to the originally grouped cases. Fig. 1 illustrates the

Table II. *Results of the discriminant analyses for classifying stroke sub-group membership according to the TUG-ABS scores (*n = 48*)*

Actual group	Stroke– Fast n (%)	Stroke– Intermediate n (%)	Stroke– Slow n (%)
Stroke-Fast (*n*=25)	22 (88)	3 (12)	0
Stroke-Intermediate (*n*=14)	1 (7.1)	11 (78.6)	2 (14.3)
Stroke-Slow (*n*=9)	0	2 (22.2)	7 (77.8)

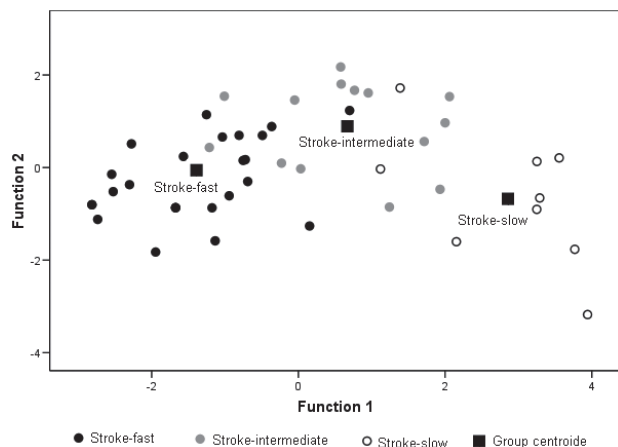TUG-ABS: Timed "Up and Go" Assessment of Biomechanical Strategies.

Fig. 1. Separation of subjects into 3 subgroups according to their Timed "Up and Go" Assessment of Biomechanical Strategies scores (*n*=48).

classifications achieved with the discriminant analyses for a better illustration of the predictive value of the TUG-ABS regarding subjects with stroke with different TUG performances.

Study 2: Construct validity by contemporary methods

Fourteen clinical PTs, who had a mean professional career experience time of 13.0 years (SD 9.6 years) (range 1.5–27 years) independently answered the semi-structured questionnaire after using the TUG-ABS with their patients. In total, 48 patients with stroke were evaluated (minimum of 3 and maximum of 4 patients for each PT).

Thirteen PTs (92.9%) evaluated the content representativeness/relevance of the set of items of the TUG-ABS as adequate. For the other questions, relevance to clinical interpretations, which could be made based upon the TUG-ABS scores, the applied utility, and the value implications of the TUG-ABS score interpretations as the basis for actions, were evaluated as adequate by all 14 PTs. Finally, all of them answered that the set of information provided by the TUG-ABS for the assessment of the biomechanical strategies during the TUG performance was also adequate. Therefore, all of the PTs judged that there were no construct under-representations, nor the presence of construct-irrelevant variance for the TUG-ABS.

Study 3: Criterion-related validity and reliability

Forty-four individuals with stroke participated, 24 men and 20 women, with a mean age of 54.7 years (SD 10.8 years) (range 30–80 years) and a mean stroke onset of 70.1 months (SD 44.5 months) (range 7–180 months). The PTs observed 2.6 (SD 0.7) TUG trials (1–4 repetitions) to verify all the TUG-ABS items by real-time observations. On the other hand, they observed 2.5 (SD 0.8) TUG trials to assess all of the items by video observations (1–4 repetitions).

Table III shows the results of the unweighted kappa statistics regarding the absolute agreement for the scores of each TUG-ABS item obtained with the video and real-time observations

(criterion-related validity). The examiners demonstrated significant and adequate absolute agreement for all items, with kappa values ranging from ($0.27 \leq \kappa \leq 0.73$; $p < 0.01$) and from ($0.29 \leq \kappa \leq 0.85$; $p < 0.005$), respectively.

Table IV provides the results of the weighted kappa statistics related to the inter-rater reliability for the individual items and the total TUG-ABS scores, obtained with real-time and video observations. For real-time observations, inter-rater reliability for the individual items and the total scores showed significant and adequate values, with kappa values in the range $0.24 \leq k \leq 1.00$; $p < 0.05$ and $k = 0.80$; $p < 0.0001$, respectively. Similar values were observed for inter-rater reliability for the items ($0.25 \leq \kappa \leq 0.94$; $p < 0.05$) and the total scores ($\kappa = 0.87$; $p < 0.0001$) by video observations.

DISCUSSION

The aim of the present study was to investigate the construct and criterion-related validities, as well as the reliability of the TUG-ABS within clinical settings. All methods employed to investigate the construct validity provided findings that supported the theoretical assumptions employed for the TUG-ABS development. In addition, agreement was observed between real-time and video observations, which illustrated the criterion-related validity of the TUG-ABS. Finally, adequate values of inter-rater reliability were found for all of the items, as well as for the total scores during both real-time and video observations.

Despite the availability of instruments for observational gait analyses of people with stroke, the majority of them are related to video-recorded performance analyses in slow motion (5, 6,

Table III. *Results of the absolute agreements according to the unweighted kappa statistics regarding the Timed "Up and Go" Assessment of Biomechanical Strategies. TUG-ABS) scores from video and real-time observations (*n = 44*)

Item	First examiner	Second examiner
Sit-to-Stand		
A	$\kappa = 0.60$; $p < 0.0001$	$\kappa = 0.53$; $p < 0.0001$
B	$\kappa = 0.73$; $p < 0.0001$	$\kappa = 0.84$; $p < 0.0001$
C	$\kappa = 0.35$; $p < 0.005$	$\kappa = 0.54$; $p < 0.0001$
Gait		
A	$\kappa = 0.37$; $p < 0.0001$	$\kappa = 0.58$; $p < 0.0001$
B	$\kappa = 0.41$; $p < 0.0005$	$\kappa = 0.45$; $p < 0.0001$
C	$\kappa = 0.44$; $p < 0.0001$	$\kappa = 0.64$; $p < 0.0001$
D	$\kappa = 0.27$; $p < 0.010$	$\kappa = 0.44$; $p < 0.0001$
E	$\kappa = 0.34$; $p < 0.005$	$\kappa = 0.35$; $p < 0.005$
Turn		
A	$\kappa = 0.54$; $p < 0.0001$	$\kappa = 0.59$; $p < 0.0001$
B	$\kappa = 0.57$; $p < 0.0001$	$\kappa = 0.69$; $p < 0.0001$
C	$\kappa = 0.43$; $p < 0.0005$	$\kappa = 0.46$; $p < 0.0005$
D	$\kappa = 0.50$; $p < 0.0005$	$\kappa = 0.29$; $p < 0.005$
Stand-to-Sit		
A	$\kappa = 0.45$; $p < 0.0001$	$\kappa = 0.39$; $p < 0.0005$
B	$\kappa = 0.36$; $p < 0.0005$	$\kappa = 0.67$; $p < 0.0001$
C	$\kappa = 0.41$; $p < 0.001$	$\kappa = 0.34$ $p < 0.0010$

Table IV. *Results of the inter-rater reliability according to the weighted kappa statistics regarding the Timed "Up and Go" Assessment of Biomechanical Strategies Scores for the real-time and video observations (n = 44)*

Item	Real-time	Video
Sit-to-Stand		
A	κ=0.49; *p*<0.0001	κ=0.58; *p*<0.0001
B	κ=1.00; *p*<0.0001	κ=0.94; *p*<0.0001
C	κ=0.65; *p*<0.0001	κ=0.38; *p*<0.0005
Gait		
A	κ=0.59; *p*<0.0001	κ=0.72; *p*<0.0001
B	κ=0.45; *p*<0.001	κ=0.15; *p*<0.050
C	κ=0.61; *p*<0.0001	κ=0.51; *p*<0.0005
D	κ=0.24; *p*<0.05	κ=0.38; *p*<0.0005
E	κ=0.58; *p*<0.0001	κ=0.27; *p*<0.010
Turn		
A	κ=0.56; *p*<0.0001	κ=0.42; *p*<0.005
B	κ=0.77; *p*<0.0001	κ=0.75; *p*<0.0001
C	κ=0.47; *p*<0.0005	κ=0.59; *p*<0.0001
D	κ=0.63; *p*<0.0001	κ=0.53; *p*<0.0001
Stand-to-Sit		
A	κ=0.60; *p*<0.0001	κ=0.49; *p*<0.0005
B	κ=0.60; *p*<0.0001	κ=0.54; *p*<0.0001
C	κ=0.38; *p*<0.0005	κ=0.25; *p*<0.001
Total score	κ=0.80; *p*<0.0001	κ=0.87; *p*<0.0001

8, 30). In addition, we are not aware of any instruments that clinically assess the biomechanical strategies adopted by people with stroke during the performance of other activities evaluated by the TUG, such as the sit-to-stand, 180°-turning, and the stand-to sit. All of these factors make comparisons between the present results with respect to the current literature difficult.

Construct validity was not reported for any of the previously cited instruments for observational gait analyses of people with stroke (5, 6, 8). Constructs are typically multidimensional and are not directly observable. Due to these factors, it is not easy to determine the construct validity of an instrument (1). Despite the difficulty in investigating the construct validity, this type of validity can be gathered by a variety of methods in an ongoing process that should address both score meaning (traditional methods) and clinical values (contemporary methods) in test interpretation and use (1, 3).

As observed in the present study, the total TUG-ABS scores were able to differentiate between people with and without stroke, and between people with stroke with different TUG times. In addition, the TUG-ABS scores were also able to correctly predict group membership for the majority of subjects with and without stroke and for the stroke subjects with fast, intermediate and fast TUG times. Fifteen healthy controls (31%) were classified as people with stroke by their scores, which indicated 15 false-positive results, whereas only one individual with stroke (2%) was classified as healthy, a false-negative. In terms of the predictive values of a tool, false-negatives are worse than false-positives, since false-positive results do not cause any harm to the subjects. On the other hand, false-negatives result in a decrease in patient care, since a patient in need of healthcare services will not receive them

(27). Therefore, all traditional methods applied for the investigation of the TUG-ABS construct validity showed adequate results, which support the theoretical assumptions behind the constructs of interest.

Positive results were also observed for the contemporary methods applied to investigate the construct validity (3). All of the PTs who used the TUG-ABS reached consensus in evaluating its construct validity. We are not aware of any study that investigated professional opinions regarding score interpretations and clinical values of previously cited observational gait tools. It is likely that the absence of construct validity investigation for those tools (5, 6, 8) is a justification for the fact that 91.8% of the PTs requested a new gait assessment tool (30). As pointed out by Toro et al. (30), "the challenge for developers of gait assessment tools is to find a balance between the practicalities of use and scientific merit". Construct validity is an important psychometric property related to the scientific merit of an instrument and, therefore, it should be further investigated.

Evidence for the practicalities of use of the TUG-ABS was found in the present study since adequate levels of absolute agreement between the analyses carried out by the real-time and video observations were observed for all the items. We are not aware of any study that has investigated the levels of absolute agreement between real-time and video observations of previously cited observational gait tools for subjects with stroke (5, 6, 8). Considering other populations, only one study was found that investigated the agreement between real-time and video observations using observational gait tools for children with disabilities and the levels of agreement were adequate for the majority of the items (31).

The item-by-item analyses showed acceptable values of reliability for the TUG-ABS, whether scored from videos or in real-time (1, 22, 27). All of the reliability values found for the TUG-ABS were similar, or even better, than those found for real-time assessments with other observational gait tools (31, 32), and for videotaped gait analysis studies (31, 33–35) and, in some cases, the videos were slowed or stopped (31, 36). These results also emphasize the feasibility of the TUG-ABS.

The TUG-ABS was developed to supplement the original TUG test, a valid, reliable and feasible measurement of basic functional mobility (14, 18, 37, 38). Therefore, the first step of the validation process could be the correlations between the TUB-ABS and the original TUG test, which was presently observed.

The TUG provides a measurement of the time spent by subjects to perform the sit-to-stand, gait, 180°-turning, and stand-to-sit in sequence, while the TUG-ABS provide a measurement of the biomechanical strategies adopted during the performance of the TUG. Time, the TUG test outcome, provides a dimension of the tasks related to performance (18). However, time alone is insufficient for diagnoses, guiding interventions, or treatment planning, since it does not allow for observation of what is impaired (39).

Besides supplementing the original TUG test, the TUG-ABS provides more detailed information, which could be used to guide professionals in their clinical decision-making. Profes-

sionals who use the TUG-ABS for clinical purposes could identify biomechanical strategies during activity performance that should be improved with future interventions. This would enable better diagnosis related to limitations in activities and allow better treatment planning. The identification of these biomechanical strategies would also indicate the need for more specific evaluations of impairments in body function and structure, and, therefore, improve the impairment diagnoses. As a standard tool, the TUG-ABS might also facilitate communication between different users and comparisons of data across time, studies and healthcare disciplines interested in functional performance and biomechanical strategies.

The present results are applicable to the characteristics of the selected subjects, which were determined in order to guarantee a wide range and the most evenly spread of variability, or heterogeneity, in their TUG performances. Therefore, future studies, with larger samples should be conducted for the investigation of the generalizability of the TUG-ABS, with various populations of subjects with stroke. The validation of a newly developed instrument is an on-going process and requires numerous research efforts, which must be achieved by complementary studies (1, 2, 10).

In conclusion, the TUG-ABS demonstrated good construct and criterion-related validities as well as adequate reliability for subjects with stroke within clinical settings, which supports the theoretical assumptions employed for its development.

REFERENCES

1. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. New Jersey: Prentice-Hall; 2009.
2. Sim J, Arnell P. Measurement validity in physical therapy. Phys Ther 1993; 73: 102–115.
3. Messick S. Validation of inferences from person's responses and performances as scientific inquiry into score meaning. Am Phsychol 1995; 50: 741–749.
4. Barak S, Duncan PW. Issues in selecting outcome measures to assess functional recovery after stroke. NeuroRX 2006; 3: 505–524.
5. Malouin F. Observational gait analysis. Gait analysis: theory and application. In: Craik R, Oatis CA, editors. Gait analysis: theory and application. St Louis: Mosby; 1995, p. 112–124.
6. McGinley JL, Goldie P, Greenwood KM, Olney SJ. Accuracy and reliability of observational gait analysis data: Judgments of push-off in gait after stroke. Phys Ther 2003; 83: 146–160.
7. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. Disabil Rehabil 2005; 27: 315–340.
8. Toro B, Nester C, Farren P. A review of observational gait assessment in clinical practice. Physiother Theory Pract 2003; 19: 137–149.
9. Faria CDCM, Teixeira-Salmela LF, Nadeau S. Development and validation of an innovative for the assessment of the biomechanical strategies: the TUG-ABS for individuals with stroke. J Rehabil Med 2013; 45: 232–240.
10. Benson J, Clark F. A guide for instrument development and validation. Am J Occup Ther 1982; 36: 789–800.
11. Davis AE. Instrument development: getting started. J Neurosci Nurs 1996; 28: 204–207.
12. Blackburn M, Van Vliet P, Mockett SP. Reliability of measures obtained with the Modified Ashworth Scale in the lower extremities of people with stroke. Phys Ther 2002; 82: 25–34.
13. Bohannon RW, Smith MB. Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. Arch Phys Med Rehabil 1997; 78: 26–32.
14. Hershkovitz A, Gottlieb D, Beloosesky Y, Brill S. Assessing the potential for functional improvement of stroke patients attending a geriatric day hospital. Arch Gerontol Geriatr 2006; 43: 243–248.
15. Centers for Disease Control and Prevention. Physical activity trends – United States, 1990–1998. Morb Mortal Wkly Rep 2001; 50: 166–169.
16. Flanbsjer U, Holmback AM, Downham D, Patten C, Lexell J. Reliability of gait performance tests in men and women with hemiparesis after stroke. J Rehabil Med 2005; 37: 75–82.
17. Janssen WGM, Bussmann HBJ, Stam HJ. Determinants of the sit-to-stand movement: a review. Phys Ther 2002; 82: 866–879.
18. Podsiadlo D, Richardson S. The Timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991; 39: 142–148.
19. Faria CDCM, Teixeira-Salmela LF, Nadeau S. Effects of the direction of turning on the Timed "Up and Go" test with stroke subjects. Top Stroke Rehabil 2009; 16: 196–206.
20. Faria CD, Teixeira-Salmela L, Silva EB, Nadeau S. Expanded Timed Up and Go Test with subjects with stroke: reliability and comparisons with matched healthy controls. Arch Phys Med Rehabil 2012; 93: 1034–1038.
21. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil 2005; 84: 719–723.
22. Sim J, Wright CC. The Kappa statistics in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005; 83: 257–268.
23. Faria CDCM, Teixeira-Salmela LF, Nadeau S. Predicting levels of basic functional mobility, as assessed by the Timed "Up and Go" test, for individuals with stroke: Discriminant analysis. Disabil Rehabil 2013; 35: 146–152.
24. Tabachnick BG, Fidell LS. Using multivariate statistics. 3rd ed. New York: HarperCollins College Publishers; 1996.
25. George SZ, Delitto A. Clinical examination variables discriminant among treatment-based classification groups: a study of construct validity in patients with acute low back pain. Phys Ther 2005; 85: 306–314.
26. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007; 30: 459–467.
27. Tooth LR, Ottenbacher KJ. The K statistic in rehabilitation research: an examination. Arch Phys Med Rehabil 2004; 85: 1371–1376.
28. Jakobsson U, Westergren A. Statistical methods for assessing agreement for ordinal data. Scand J Caring Sci 2005; 19: 427–431.
29. Vieira AJ, Garrett JM. Understanding inter-observer agreement: the kappa statistic. Family Med 2005; 37: 360–363.
30. Toro B, Nester CJ, Farren PC. The status of gait assessment among physiotherapists in the United Kingdon. Arch Phys Med Rehabil 2003; 84: 1878–1884.
31. Wren TAL, Rethlefsen SA, Healy BS, Do KP, Dennis SW, Kay RM. Reliability and validity of visual assessments of gait using a modified physician rating scale for crouch and foot contact. J Pediatr Orthop 2005; 25: 646–650.

32. Kawamura CM, Morais Filho MC, Barreto MM, Asa SKP, Juliano Y, Novo NF. Comparison between visual and three-dimensional gait analysis in patients with spastic diplegic cerebral palsy. Gait Posture 2007; 25: 18–24.

33. Stott NS, Atherton WG, Mackey AH, Galley IJ, Nicol RO, Wash SJ. Reliability and validity of assessment of sagittal plane deviations in children who have spastic diplegia. Arch Phys Med Rehabil 2005; 86: 2337–2341.

34. Keenan AM, Bach TM. Video assessment of rearfoot movements during walking: a reliability study. Arch Phys Med Rehabil 1996; 77: 651–655.

35. Mackey AH, Lobb GL, Walt SE, Stott NS. Reliability and validity of the Observational Gait Scale in children with spastic diplegia. Dev Med Child Neurol 2003; 45: 4–11.

36. Eastlack ME, Arvidson J, Snyder-Macler L, Danoff JV, McGarvey CL. Interrater reliability of videotaped observational gait-analysis assessments. Phys Ther 1991; 71: 465–472.

37. Ng SS, Hui-Chan CW. The timed up & go test: its reliability and association with lower-limb impairments and locomotor capacities in people with chronic stroke. Arch Phys Med Rehabil 2005; 86: 1641–1647.

38. Salbach NM, Mayo NE, Higgins J, Ahmed S, Finch LE, Richards CL. Responsiveness and predictability of gait speed and other disability measures in acute stroke. Arch Phys Med Rehabil 2001; 82: 1204–1212.

39. Fisher WP, Harvey RF, Taylor P, Kilgore KM, Kelly CK. Rehabits: a common language of functional assessment. Arch Phys Med Rehabil 1995; 76: 113–122.

APPENDIX I. Semi-structured questionnaire for the investigation of the construct validity

Number of subjects with stroke, who were evaluated with the Timed "Up and Go" Assessment of Biomechanical Strategies (TUG-ABS) _____

The questions below were formulated based upon the different aspects of construct validity proposed by Messick in the following publication: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *Am Phsychol.* 1995; 50: 741–749. **For each of the questions, please check the category response that you judge most appropriate:**

A. Is the content of the set of items of the TUG-ABS representative/relevant to the domain of interest (biomechanical strategies exhibited by individuals with hemiparesis due to stroke during their TUG performances)?

    (1) Not representative/relevant
    (2) Somewhat representative/relevant
    (3) Quite representative/relevant
    (4) Highly representative/relevant

B. Is the set of items of the TUG-ABS relevant to the clinical interpretations, which can be made based upon these measures?

    (1) Not relevant
    (2) Somewhat relevant
    (3) Quite relevant
    (4) Highly relevant

C. Please, evaluate the value of the use (applied utility) of the TUG-ABS:

    (1) Not valuable
    (2) Somewhat valuable
    (3) Quite valuable
    (4) Highly valuable

D. Please, evaluate the value implications of the TUG-ABS score interpretations as a basis for actions:

    (1) Not relevant
    (2) Somewhat relevant
    (3) Quite relevant
    (4) Highly relevant

E. In your opinion, does the TUG-ABS provides enough information for the assessment of the biomechanical strategies during the TUG performance, that is, it is:

    (1) Too narrow and fails to include important dimensions or facets of the construct" (presence of construct underrepresentation)
    (2) Too broad, contains excess of variance associated with other distinct constructs, as well as method variance such as response sets or guessing propensities which affects responses in a manner irrelevant to the interpreted construct"(presence of construct-irrelevant variance)
    (3) Adequate, since it includes only dimensions or facets which are strictly associated and affect responses in a manner relevant to the interpreted construct