J Rehabil Med 2013; 45: 232–240

ORIGINAL REPORT

DEVELOPMENT AND VALIDATION 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 develop and validate a clinical tool based on the biomechanical strategies exhibited by people with hemiparesis due to stroke during the performance of the Timed "Up and Go" test.**

Design/methods: **The Timed "Up and Go" Assessment of Biomechanical Strategies (TUG-ABS) was developed for subjects with stroke, based on the analyses of 3 sources of information: published evidence; opinions of rehabilitation professionals; and observations of TUG performances, followed by a multi-step approach, which involved the investigation of the reliability, content, and criterion-related validity of the preliminary version. Content validity was established by an expert panel, whereas intra- and inter-rater reliability was established by two independent examiners. Criterion-related validity was established by comparing the TUG-ABS scores at the item level obtained by independent analyses of video observations and the gold standard motion analysis system. The final tool included the items, which showed acceptable values for these psychometric properties.**

Results: **The preliminary version consisted of 24 items with 3 response categories. Twenty-one items showed acceptable content validity ($0.72 \leq \kappa \leq 1.00$; $p \leq 0.01$), 19 acceptable intra- and inter-rater reliability ($0.36 \leq \kappa \leq 1.00$; $p \leq 0.04$), and 15 acceptable criterion-related validity ($0.29 \leq \kappa \leq 1.00$; $p \leq 0.04$).**

Conclusion: **The final developed 15-item TUG-ABS version proved to be valid and reliable for individuals with hemiparesis due to stroke, but it should be clinically validated before being used for clinical applications and research purposes.**

Key words: stroke; assessment; mobility; instrument development; reliability; content validity.

J Rehabil Med 2013; 45: 232–240

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

Accepted Oct 9, 2012; Epub ahead of print Feb 6, 2013

INTRODUCTION

Stroke remains one of the most common causes of long-term disability worldwide and is a major public health concern (1). Thus, the understanding of the disabilities following stroke is a high priority for healthcare systems. Among all of the common disabilities of stroke survivors, those related to functional mobility are the most prevalent and disabling (2), and are the primary targets for rehabilitation.

A widely recommended test to assess functional mobility after stroke (3–7) is the Timed "Up and Go" (TUG) (6), proposed as a modified timed version of the "Get-up and Go" test (8). The TUG has shown high levels of validity to assess functional mobility (6), is highly reliable (3, 5) and responsive (3, 4) for individuals with stroke, and is able to differentiate between people with stroke from the healthy elderly population (5) and people with stroke with various degrees of disabilities (4). It can also be used as a predictive measure of disabilities (4). Finally, the TUG is easily administered, requires no training or specialized equipment, and can be used in both community and institutional settings (7).

Despite all of the advantages of the TUG to assess the functional mobility of people with stroke, the only investigated outcome of this test is the time spent to perform the sequential activities (3–7). Although time provides a dimension of the tasks related to performance (6), as stated by Fisher et al. (9), time alone is insufficient for guiding interventions, treatment planning, or diagnoses and does not allow the functional construct be articulated and observed. Therefore, the TUG does not provide clinicians with sufficient information on the specific forms of movement dysfunctions that a subject may demonstrate.

Biomechanical strategies have been widely used by rehabilitation professionals to analyse performance and guide clinical decision-making. In addition, important changes in biomechanical strategies exhibited by individuals with stroke during the performance of activities evaluated by the TUG are

well established and widely employed for the understanding of stroke disabilities (10–13). Therefore, it is necessary to develop a clinical measure for the systematic evaluation of the biomechanical strategies during the TUG performance. This could provide clinicians with additional and relevant information on which to base clinical decision-making, in a feasible and systematic way, and with acceptable values of reliability and validity.

Despite the availability of some tools for observational gait analyses of people with stroke, one of the TUG activities, previous studies have highlighted some important limitations of their developmental processes and psychometric properties (14–17). In the clinical context, there is also a need for an adequate clinical gait assessment tool (16). Also, no tools exist regarding the clinical evaluations of the biomechanical strategies exhibited by people with stroke during the performance of all of the other TUG activities. Therefore, the aim of this study was to develop and validate a clinical tool based on the biomechanical strategies exhibited by individuals with stroke during their TUG performances: the TUG – Assessment of Biomechanical Strategies (TUG-ABS).

METHODS

The systematic processes of the development of the TUG-ABS followed previous guidelines and recommendations (18–20). First, the target population for the use of the TUG-ABS was determined: people with stroke over 20 years of age from the general community, with motor impairments, as characterized by residual weaknesses and/or increased tonus of the paretic lower limb (21, 22); with the ability to follow instructions; and able to perform the TUG with or without assistive devices.

Subsequently, the characteristics of the variables to be constituted as TUG-ABS items were determined. These variables were related to the biomechanical strategies adopted by subjects with stroke during the performance of the activities evaluated by the TUG. In addition, they should be able to differentiate between individuals with and without stroke and between individuals with stroke with various levels of TUG performances, since the TUG times were able to differentiate between subjects with stroke with mild, moderate, and severe neurological impairments (4). In addition, subjects with stroke could be divided into 3 sub-groups regarding their TUG times: fast, moderate and slow (23). Finally, these variables should also be easily observed within the clinical contexts, i.e., with video or real-time observations.

In line with these guidelines (18–20), the preliminary version of the TUG-ABS was developed based on the analyses of 3 different sources of information: published evidence; opinions of rehabilitation professionals; and observation of TUG performances. The sampling of studies to be reviewed and participants (professionals or patients) was based on sampling to redundancy, the criterion recommended for the process of development of the preliminary version of an instrument (20).

A comprehensive and critical literature review was carried out to identify the variables of interest and to guide the selection of both item format and characteristics, as well as the creation of the scoring model. In addition, physical therapists (PT) involved in stroke rehabilitation and, therefore, who could be assigned as the target group to use the TUG-ABS (18, 19), replied to a written-structured questionnaire. They provided their opinions regarding the most relevant and essential variables, which could be selected to constitute the TUG-ABS items (18). Furthermore, a database with previously recorded videos with TUG performances of subjects with stroke (24) was used for analyses, employing video resources to select the variables that could constitute the items of the tool. A database with previously recorded videos with TUG performances of healthy subjects, matched by age, gender, and levels of physical activity (24) was also analysed to allow for the identification of variables that could differentiate between subjects with and without stroke. The videos of the subjects with stroke were grouped according to their levels of TUG performances (fast, moderate, and slow) (23) to allow for the identification of variables that could differentiate between individuals with stroke with various performance levels.

According to previous instructions (18, 20), the 3 sources of information were used in a complementary manner, to provide direction for the instrument development and to validate information obtained from all sources. All of the information obtained was extensively and systematically analysed. As previously recommended (18, 19), the number of items initially developed for the preliminary version should exceed the desired final length by 1.5–2.0 times, to ensure that there were a sufficient number of items in the pool after testing. Therefore, the preliminary version of the TUG-ABS was developed with 24 items: 5 related to the sit-to-stand, 8 to gait, 5 to turning, and 6 to the stand-to-sit (Table I). Considering the item scaling, the scoring model selected for all of the developed items was polytomous, on an ordinal scale of measurement, with 3 possible response categories, ranging between 1 and 3, since the tool was intended to differentiate between individuals with stroke with different levels of TUG performances.

After the development of the preliminary version of the TUG-ABS, a multi-step approach was applied to establish its final version, which involved the investigation of its content validity (first phase), intra- and inter-rater reliability (second phase), and criterion-related validity (third phase). As previously recommended (18, 19), the product of this multi-step approach resulted in the final tool comprised of the items, which showed acceptable values for these 3 psychometric properties.

First phase: content validation of the TUG-ABS

Content validity was investigated by an expert panel, which was composed of internationally well-known PTs involved in stroke rehabilitation and who had multiple publications in refereed journals and conference proceedings. All of the procedures followed the specific recommendations regarding the selection and the use of the content of the experts (26), experts' judgments in content-related validity evidence (27), and the determination and quantification of the content validity (28).

After the analyses of the preliminary version of the TUG-ABS, the experts judged the consistency of the conceptual definitions, the representativeness and the relevance in the domain of interest, clinical interpretations, clarity and understanding of each item, and of the nature of the overall instrument (26, 27). All of these characteristics were evaluated on a 4-point ordinal relevance or representative rating scale (1 = not relevant/representative; 2 = somewhat relevant/representative; 3 = quite relevant/representative; 4 = highly relevant/representative) (26, 28). The experts were also asked to make any suggestions or modifications for the adequacy of each item, as well as for the addition or exclusion of items. The following recommendation made by Polit et al. (28) was also adopted: "Unless only minor item revisions are needed based on the first round results, a second round of expert review should be conducted", which should follow the same procedures.

The content validity was investigated according to traditional subjective processes (26, 27), which involved the discussion and consideration of any suggestions provided by the experts, and the quantitative content validity index (CVI) at the item level, which involved statistical analyses. As proposed by Polit et al. (28), the CVI was linked to the modified kappa-like index, which adjusts for the chance of agreement regarding relevance. The CVI was computed by the number of experts, who rated the items either at 3 or 4, divided by the total number of experts, for the proportion of agreement regarding relevance. The modified kappa-like index was calculated based on the probability of a chance occurrence, using the formula for a binominal random variable and the kappa designating agreement of relevance. To determine if the obtained CVI values were adequate, the table provided by Polit et al.

Table I. *Original 24 items in the preliminary version of the Timed "Up and Go" Assessment of Biomechanical Strategies and the added items recommended after the first round of expert panel analyses*

Original items
Sit-to-stand
A) Support of the upper limb(s) associated with lateral trunk flexion and/or trunk rotation
B) Attempts used to perform the sit-to-stand associated with the strategy of approaching the pelvis to the extremity of the seat
C) Momentum generated by the first anterior trunk flexion and extension of the trunk and lower limbs
D) Lateral trunk deviation
E) Transition between the sit-to-stand and gait
Gait
A) Step symmetry and step length
B) Initial contact with the heel
C) Hip extension during the stance phase
D) Swing phase – foot clearance off the ground
E) Step length considering the relationship between the foot of the initial contact and the foot of support
F) Displacement and/or trunk oscillation in the three planes of movement
G) Weight bearing on the lower limbs
H) Use of plantar flexion during the push-off
Turn
A) Relationships between the outer and inner foot during the turn
B) Steps used only to perform the turn. (Do not consider the steps used for the gait immediately before and/or after the turn)
C) Trajectory of the advancing lower limb
D) Sequence of gait-turning-gait
E) Turning of the body – trajectory of the outward shoulder
Added items
Sit-to-stand
New item Y) Feet positioning and active knee flexion
Gait
New item Y) Lower limbs forward progression
New item Z) Sign of weight bearing hesitation on the lower limbs
Turn
New item Y) Trunk rotation in the reversal of direction during the turn
New item Z) Displacement of the advancing lower limb in the reversal of direction during the turn

(28) with the values of adequate modified kappa-like index regarding the number of participant experts was also employed.

Second phase: intra- and inter-rater reliability of the TUG-ABS

The intra- and inter-rater reliabilites of the TUG-ABS were investigated for the pool of items that reached adequate CVI values linked to the modified kappa-like index levels.

For this phase, individuals with stroke with the previously described characteristics were recruited. The sample size was determined to guarantee a minimum number of subjects, which provided a wide range and the most even spread of variability in TUG performances, as was recommended (18, 19, 29) and previously adopted (15, 30–32).

Before data collection, eligible participants were informed of the objectives of the study and were asked to provide consent, which was approved by the university research ethics review board. Demographic and clinical data were collected for all subjects by the same PT.

To perform the TUG, the subjects sat in a chair (depth 45 cm, width 49 cm, arm rest height 20 cm) (3), whose height was adjusted to 100% of their leg length and the back rest adjusted to a trunk position of 90º (7, 33). Subjects 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-m mark, turned around, walked back, and sat down in the chair (6). After a familiarization trial, the TUG was performed (34). If there was a risk of falling, the examiner followed the subjects half a step behind, so as to not influence their walking pace (3, 10).

Three video cameras (JVC®, GR-DVL 9800) were used to record the TUG performances. They were positioned in the frontal plane, and left and right sagittal planes and had acquisition frequencies of 30 Hz. Only one TUG performance was recorded for each subject and the video was processed and edited by Adobe® After Effects CS3® software, which allowed grouping the 3 views into the same file (24). This means that all of the 3 views could be observed simultaneously on a single screen.

The recorded videos were analysed by two independent examiners after a period of familiarization with the TUG-ABS. They observed the videos at normal speeds, without stopping or slowing movements, as many times as necessary to score all of the items. However, they were instructed to try to score all items with minimum video repetitions. Although the TUG-ABS should be administered by real-time observations, the recorded videos were used in this study to avoid biases related to changes in the subjects' performances, which could have occurred in real-time observations (29, 35).

For the intra-rater reliability, the examiners rated identical second evaluation sessions (test 2), 4 weeks apart. No feedback or further discussion was allowed between the two time-points. To avoid biases related to memory, the subjects' face was pixelated using Adobe® After Effects CS3® (24) and the videos were shown randomly for each observation session.

The levels of agreement between raters and evaluations were obtained according to the quadratic weighted kappa statistics (36) ($\alpha = 5\%$). If the levels of agreement were beyond those expected by chance, and therefore reached significance level, the kappa values were interpreted as follows: 0.01–0.020, 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 or almost perfect levels of agreement (29, 30, 36–38).

Third phase: criterion-related validity of the TUG-ABS

For performance-based tests related to observational analyses like the TUG-ABS, the recommended method to determine criterion-related validity is the comparison of the measures obtained by one examiner using observational analyses with those obtained by another examiner using quantitative data provided by computerized motion analysis systems (gold standard) (15–17, 35, 39–41). Both video and motion analysis system sets of data should be recorded at the same time. Therefore, examiner 1 scored the TUG-ABS items by observing the recorded videos, while examiner 2 analysed the quantitative data provided by the motion analysis system. These examiners were not involved in any of the previous procedures, nor in the data collection and processing. No feedback or discussion was allowed between the examiners. In addition, to perform these analyses, clear and objective criteria were established and employed by both examiners for each item of the TUG-ABS, based on published information regarding the biomechanical strategies during the sit-to-stand (12, 42, 43), gait (11, 43–46), 180º-turning (10, 47, 48), and stand-to-sit (12, 42, 43). In this manner, the criteria and target ratings were independent and free from biases (35).

To investigate the criterion-related validity of the TUG-ABS, individuals with stroke with the previously described characteristics and who had not participated in the previous phase, were recruited. All of the previously described procedures regarding demographic, clinical data collection, and TUG performances were followed. During the TUG performance, data collection was simultaneously obtained by the computerized motion analysis system and 3 video cameras (JVC®, GR-DVL 9800), which were synchronized.

The motion analysis system consisted of 4 camera units of the Optotrak® system (Northern Digital Technology, Waterloo, ON, Canada) (44) and 3 AMTI® force plates (Advanced Mechanical Technologies, Watertown, MA, USA). An adjustable instrumented chair with 4 AMTI strain gauge transducers (MC3A-3-250, AMTI, Newton, MA, USA) was also used to record orthogonal forces under each thigh (49). The

AMTI® force plates were embedded in the 10-m walkway and data was acquired at 600 Hz. The seat height of the instrumented chair was adjusted for each individual's leg length (12), and data were also collected at 600 Hz. The markers were placed on the feet, shins, thighs, pelvis, and trunk. Seventeen anatomical landmarks were also digitized with a 6-marker probe to precisely define the segments and the anatomical axes (44). The 3D coordinates were recorded at a sampling frequency of 60 Hz, as previously carried out (44) and recommended (40). After data collection, kinematic data were filtered with a 4th-order Butterworth zero-lag filter, with a cut-off frequency of 6 Hz, and the relative angles using a Cardanic (x–y–z) rotation sequence (44). The force plate data were filtered with a 4th-order Butterworth zero-lag filter with a cut-off frequency of 10 Hz, and were re-sampled at 60 Hz to match the kinematic data. An inverse dynamic approach was used to estimate the net joint moments and net powers (41, 44).

The video cameras had acquisition frequencies of 30 Hz and the recorded videos were processed and edited as previously described. Although the videos were recorded at 30 Hz, they were analysed in real time, without stopping or slowing speed motion. From a technical standpoint, any dynamic event happening faster than 83 ms could not be perceived by the human eye (50, 51). On the other hand, both the kinematic and kinetic data were sampled at 60 Hz (16.7 ms). Therefore, a valid event was determined, by examiner 2, as a set of 5 or more frames (≥ 83.5 ms).

A total of 3–5 TUG trials were collected to ensure sufficient data for analyses. However, only one trial for each subject was selected, considering the variability in the TUG performances, which allowed the observation of all of the response categories for each item of the TUG-ABS.

The levels of absolute agreement between the data of the observational analyses and the motion analysis system for each item of the TUG-ABS were rated according to the unweighted kappa statistics (36), to investigate the absolute agreement for categorical polytomous data ($\alpha = 5\%$). If the levels of agreement were beyond those expected by chance, and therefore reached significance levels, the kappa values were interpreted as previously described.

RESULTS

First phase: content validation of the TUG-ABS

The content validity of the preliminary version of the TUG-ABS (Table I) was judged by 8 selected experts, who had a mean career duration in the areas of biomechanics and rehabilitation of people with stroke of 13.4 years (standard deviation (SD) 5.6; range 7–22 years). Two-round analyses for the establishment of the content validity were performed by the experts, totalling a mean analysis time of 236 min (SD 109; range 120–420 min), with a total of 32 h of analysis time. Of the 24 items evaluated in the first round (Table I), 7 did not reach acceptable CVI values linked to the modified kappa-like index levels (< 0.72), and, therefore, were excluded (Tables II–V). The expert panel recommended the addition of 5 new items (Table I).

During the second round, two of the new items did not reach acceptable CVI values linked to the modified kappa-like index levels, and, therefore, were excluded (Tables II–V). All of the other 20 items reached acceptable values (> 0.72), and, therefore, were retained. In addition, consensus was established by the expert panel regarding their consistency with the conceptual definitions, representativeness, and relevance to the domain of interest, relevance to clinical interpretations, and clarity and comprehensiveness of the 20 items, with modified kappa-like index values ranging from 0.72 to 1.00 (Tables II–V). Furthermore, all of the experts evaluated the set of items for each activity and for the whole instrument as sufficient to represent the content ($\kappa = 1.00$).

Table II. *Content validity, reliability, and criterion-related validity of the sit-to-stand items of the Timed "Up and Go" Assessment of Biomechanical Strategies (TUG-ABC)*

Measurement property	Statistics	Item A	Item B	Item C	Item D	Item E	Item Y
Content validity							
Consistency with conceptual definitions	CVI	1.00	1.00	1.00	1.00	0.75	0.75
	κ	1.00	1.00	1.00	1.00	0.72	0.72
Representativeness/ relevance to the domains	CVI	1.00	1.00	1.00	1.00	0.63[a]	0.75
	κ	1.00	1.00	1.00	1.00	0.52[a]	0.72
Relevance to clinical interpretations	CVI	1.00	0.88	1.00	1.00	0.63[a]	0.75
	κ	1.00	0.88	1.00	1.00	0.52[a]	0.72
Clarity and comprehensiveness	CVI	1.00	1.00	1.00	1.00	0.75	1.00
	κ	1.00	1.00	1.00	1.00	0.72	1.00
Reliability							
Intra-examiner 1	κ	0.67	0.87	0.95	0.68	–	1.00
	p	< 0.01	< 0.01	< 0.01	< 0.01		< 0.01
Intra-examiner 2	κ	0.95	0.94	1.00	1.00	–	0.39
	p	< 0.01	< 0.01	< 0.01	< 0.01		< 0.05
Inter-examiner first evaluation	κ	0.90	0.87	0.91	0.62	–	0.08
	p	< 0.01	< 0.01	< 0.01	< 0.01		0.37*
Inter-examiner second evaluation	κ	0.90	0.94	0.95	0.81	–	0.11
	p	< 0.01	< 0.01	< 0.01	< 0.01		0.35*
Criterion-related validity	κ	0.86	1.00	0.75	0.09	–	–
	p	< 0.001	< 0.001	< 0.001	0.126*		
Final decision related to the final version of the TUG-ABS		Included as Item A	Included as Item B	Included as Item C	Excluded	Excluded	Excluded

*$p \geq 0.05$.
[a]CVI values with modified kappa-like index < 0.72.
CVI: content validity index.

Table III. *Content validity, reliability, and criterion-related validity of the gait items of the Timed "Up and Go" Assessment of Biomechanical Strategies (TUG-ABS)*

Measurement property	Statistics	Item A	Item B	Item C	Item D	Item E	Item F	Item G	Item H	Item Y	Item Z
Content validity											
Consistency with	CVI	1.00	1.00	1.00	1.00	0.63[a]	0.75	0.75	1.00	0.88	0.88
conceptual definitions	κ	1.00	1.00	1.00	1.00	0.52[a]	0.72	0.72	1.00	0.88	0.88
Representativeness/	CVI	1.00	0.88	1.00	1.00	0.50[a]	0.5[a]	0.63[a]	1.00	1.00	0.88
relevance to the domain	κ	1.00	0.88	1.00	1.00	0.31[a]	0.31[a]	0.52[a]	1.00	1.00	0.88
Relevance to clinical	CVI	0.88	0.88	1.00	0.88	0.5[a]	0.38[a]	0.38[a]	1.00	0.88	0.63[a]
interpretations	κ	0.88	0.88	1.00	0.88	0.31[a]	0.20[a]	0.20[a]	1.00	0.88	0.52[a]
Clarity and	CVI	1.00	1.00	1.00	1.00	0.75	0.88	0.50[a]	1.00	0.88	0.88
comprehensiveness	κ	1.00	1.00	1.00	1.00	0.72	0.88	0.31[a]	1.00	0.88	0.88
Reliability											
Intra-examiner 1	κ	0.76	0.90	0.71	0.76	–	–	–	0.74	0.86	–
	p	<0.001	0.001	0.005	<0.001				0.005	<0.001	
Intra-examiner 2	κ	0.94	1.00	1.00	0.94	–	–	–	1.00	1.00	–
	p	<0.001	<0.001	<0.005	<0.001				<0.001	<0.001	
Inter-examiner first	κ	0.70	0.80	0.60	0.70	–	–	–	0.61	0.45	–
evaluation	*p*	<0.005	<0.01	<0.05	<0.005				<0.01	<0.05	
Inter-examiner second	κ	0.88	0.90	0.70	0.88	–	–	–	0.91	0.56	–
evaluation	*p*	<0.001	<0.001	<0.01	<0.001				<0.001	<0.01	
Criterion-related validity	κ	0.83	0.67	0.50	0.79	–	–	–	0.14	0.75	–
	p	<0.001	<0.005	<0.05	<0.005				0.053*	<0.001	
Final decision related to the final version of the TUG-ABS		Included as Item A	Included as Item B	Included as item C	Included as item D	Excluded	Excluded	Excluded	Excluded	Included as item E	Excluded

*$p \geq 0.05$.
[a]CVI values with modified kappa-like index <0.72.
CVI: content validity index.

Second phase: intra- and inter-rater reliability of the TUG-ABS
A total of 22 individuals with stroke participated, 12 men and 10 women, mean age of 54.7 years (SD 15.4; range 26–80 years) and mean time since onset of stroke 52.2 months (SD 49.2; range 2–155 months). The examiners observed a mean of 4.8 video repetitions (SD 1.1; range 3–7 video repetitions) of the TUG trials in the first and 4.0 repetitions (SD 3.5; range 2–8 repetitions) in the second session, to score all the TUG-ABS items.

The majority of the items that reached acceptable content validity, showed acceptable inter- and intra-rater reliability, with

Table IV. *Content validity, reliability, and criterion-related validity of the turn items of the Timed "Up and Go" Assessment of Biomechanical Strategies (TUG-ABS)*

Measurement property	Statistics	Item A	Item B	Item C	Item D	Item E	Item Y	Item Z
Content validity								
Consistency with conceptual definitions	CVI	0.88	1.00	0.75	1.00	0.75	0.88	0.63[a]
	κ	0.88	1.00	0.72	1.00	0.72	0.88	0.52[a]
Representativeness/relevance to the domains	CVI	0.88	1.00	0.75	1.00	0.63[a]	0.88	0.63[a]
	κ	0.88	1.00	0.72	1.00	0.52[a]	0.88	0.52[a]
Relevance to clinical interpretations	CVI	1.00	1.00	0.75	1.00	0.63[a]	0.75	0.50[a]
	κ	1.00	1.00	0.72	1.00	0.52[a]	0.72	0.31[a]
Clarity and comprehensiveness	CVI	1.00	1.00	0.63[a]	1.00	0.88	0.75	0.63[a]
	κ	1.00	1.00	0.52[a]	1.00	0.88	0.72	0.52[a]
Reliability								
Intra-examiner 1	κ	0.89	0.77	–	0.80	–	0.86	–
	p	<0.001	<0.001		<0.005		<0.001	
Intra-examiner 2	κ	1.00	1.00	–	0.89	–	1.00	–
	p	<0.005	<0.001		<0.001		<0.001	
Inter-examiner first evaluation	κ	0.68	0.73	–	0.57	–	0.75	–
	p	<0.01	<0.005		<0.01		<0.005	
Inter-examiner second evaluation	κ	0.68	0.71	–	0.86	–	0.59	–
	p	<0.01	<0.005		<0.001		<0.01	
Criterion-related validity	κ	0.87	0.29	–	0.49	–	0.50	–
	p	<0.001	<0.05		<0.01		<0.01	
Final decision related to the final version of the TUG-ABS		Included as item A	Included as item B	Excluded	Included as item D	Excluded	Included as item C	Excluded

[a]CVI values with modified kappa-like index <0.72.
CVI: content validity index.

Table V. *Content validity, reliability, and criterion-related validity of the stand-to-sit items of the Timed "Up and Go" Assessment of Biomechanical Strategies (TUG-ABS)*

Measurement property	Statistics	Item A	Item B	Item C	Item D	Item E	Item F
Content validity							
Consistency with conceptual definitions	CVI	0.88	1.00	0.63[a]	1.00	1.00	1.00
	κ	0.88	1.00	0.52[a]	1.00	1.00	1.00
Representativeness/relevance to the domains	CVI	1.00	1.00	0.50[a]	1.00	1.00	1.00
	κ	1.00	1.00	0.31[a]	1.00	1.00	1.00
Relevance to clinical interpretations	CVI	1.00	1.00	0.50[a]	1.00	1.00	1.00
	κ	1.00	1.00	0.31[a]	1.00	1.00	1.00
Clarity and comprehensiveness	CVI	1.00	0.88	0.75	1.00	1.00	1.00
	κ	1.00	0.88	0.72	1.00	1.00	1.00
Reliability							
Intra-examiner 1	κ	0.79	0.95	–	0.01	0.83	0.90
	p	<0.005	<0.001		0.41*	<0.001	<0.001
Intra-examiner 2	κ	0.88	1.00	–	1.00	0.86	0.89
	p	<0.001	<0.001		<0.001	<0.001	<0.001
Inter-examiner first evaluation	κ	0.55	0.77	–	0.25	0.77	0.61
	p	<0.01	<0.005		0.16*	<0.005	<0.05
Inter-examiner second evaluation	κ	0.53	0.82	–	0.05	0.76	0.61
	p	<0.01	<0.005	–	0.59*	<0.005	<0.05
Criterion-related validity	κ		0.57	–	–	0.23	0.61
	p		<0.01	–	–	0.10*	<0.001
Final decision related to the final version of the TUG-ABS		Included as item A	Included as item B	Excluded	Excluded	Excluded	Included as item C

*$p \geq 0.05$.
[a]CVI values with modified kappa-like index <0.72.
CVI: content validity index.

kappa coefficient values ranging from 0.36 to 1.00 ($p \leq 0.04$) (Tables II–V). Only two items (Tables II and V) did not show levels of agreement beyond those expected by chance and, therefore, were excluded.

Third phase: criterion-related validity of the TUG-ABS

Thirteen individuals with stroke participated, 6 men and 7 women, with a mean age of 63.4 years (SD 13.1; range 46–86 years) and a mean time since the onset of their stroke of 79.9 months (SD 32.7; range 44–160 months). The examiner observed a mean of 3.9 video repetitions (SD 0.6; range 3–5) of the TUG trials to score all the TUG-ABS items.

Item D of the sit-to-stand ($\kappa = 0.09$; $p = 0.126$) (Table II), H of gait ($\kappa = 0.14$; $p = 0.053$) (Table III), and E of the stand-to-sit ($\kappa = 0.23$; $p = 0.097$) (Table V) did not reach significant levels of agreement and, therefore, were excluded. All of the other 15 items showed acceptable absolute agreement values ($0.29 \leq \kappa \leq 1.00$; $p \leq 0.037$) (Tables II–V).

Therefore, the final version of the TUG-ABS was established with a total of 15 items, 3 related to sit-to-stand, 5 related to gait, 4 related to turn, and 3 related to stand-to-sit (Appendix I).

DISCUSSION

After the systematic, clear, and objective processes of the development of the TUG-ABS, followed by the multi-step approach to investigate its psychometric properties, the final version of the instrument was established. The 15-item TUG-ABS proved valid and reliable for subjects with stroke. As pointed out by Benson & Clark (18), when the planning, construction, and content validation investigation of a new instrument are properly undertaken, the other steps related to its validation are easily accomplished, which were observed for the TUG-ABS.

For all the previously cited observational gait tools, content validation analyses were not reported (14, 15, 17). The confusion between content and face validities, and the previously unquantified nature of content validity, have led to misunderstandings of the importance of this psychometric property (52). As stated by Lynn (52), "Content validity, by its nature and definition, demands assessment rigor, and its assessment is critical". In the present study, rigorous content validation processes were carried out and quantified, followed by statistical analyses. All of the items included in the final version of the TUG-ABS showed evidence of adequate content validity (28) for all investigated characteristics.

In addition, all of the items of the final TUG-ABS version showed kappa values greater than 0.40 and, therefore, were classified as moderate, good or very good levels of intra- and inter-rater reliability, where 66.7% were classified as good or very good ($\kappa > 0.60$). These values were greater than those reported by videotaped gait analysis studies, which reported intra- and inter-rater values of $0.45 \leq \kappa \leq 0.49$ (53), $\kappa = 0.19$ (54), $0.30 \leq \kappa \leq 0.69$ (38), and $0.29 \leq \kappa \leq 1.00$ (37) and even, in some cases, the videos were slowed or stopped: $0.11 \leq \kappa \leq 0.52$ (30) and $0.36 \leq \kappa \leq 0.74$ (38).

Two previous reviews (15, 17) reported that studies investigating criterion-related validity using gold standard measures and employing adequate criteria, with valid and objective

categories, and appropriate statistical analyses were lacking. Those factors seriously compromised the justification of the use of these instruments. In the present study, concurrent criterion-related validity was established for all of the items included in the final version of the TUG-ABS. Of the 15 items, 14 (93.3%) showed absolute agreement values for concurrent criterion-validity, as classified as moderate, good or very good ($\kappa \geq 0.47$). Only item B on the turn, showed fair agreement value. These values were similar to those reported by the only previous study found that also investigated the concurrent criterion-validity at the item level for observational gait analyses that applied quantitative kinematic and kinetic data and kappa statistics for absolute agreements: $0.38 \leq \kappa \leq 0.94$, which were considered as high validity values (37).

Despite the positive results of the present study, some limitations should be considered. First of all, it is possible that some important variables regarding the biomechanical strategies of the TUG performances of subjects with stroke were not included in the TUG-ABS, despite the rigorous processes involved in the planning and development of the instrument, as well as the positive results observed in its content validation. Therefore, more investigations and clinical applications of this tool are necessary to determine if the selected variables are adequate to allow a comprehensive evaluation of the biomechanical strategies of subjects with stroke. Another limitation was the use of videos of the subjects' TUG performances, which was used to limit the bias in scoring the TUG-ABS, consistent with the methodological rigour of this study. As a clinically-oriented instrument, it is necessary that future studies determine if video analyses scores of the TUG-ABS are similar to those with real time observations.

Finally, the present results are restricted to the characteristics of the selected subjects, which were determined 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 should be achieved by complementary studies (18, 35, 55).

In conclusion, the final developed 15-item TUG-ABS version proved to be valid and reliable for individuals with stroke, but still needs to be clinically validated, before being used for clinical applications and research purposes.

ACKNOWLEDGEMENTS

Financial support was provided by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Graduate Student's Exchange Program (Government of Canada), Student Dissertation Award of the International Society of Biomechanics (ISB), the Canadian Institutes of Health Research (CIHR) and REPAR. S. Nadeau is a senior researcher supported by Fonds de la Recherche en Santé du Québec. The authors are also grateful to Dr John Henry Salmela for copy-editing the manuscript, and to Cyril Duclos, Pierre Desjardins, Daniel Marineau, Michel Goyette, Fernando de Aguiar Faria and Abraão Marques Tavares for their technical support.

The authors are also grateful to Dr. John Henry Salmela and Dr. Louise Ada for copy-editing the manuscript.

REFERENCES

1. Murray CJL, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global burden of disease study. Lancet 1997; 349: 1436–1442.
2. Geyh S, Cieza A, Schouten J, Dickson H, Frommelt P, Omar Z, et al. ICF Core Sets for stroke. J Rehabil Med 2004; Suppl 44: 135–141.
3. 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.
4. 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.
5. 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.
6. 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.
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. Mathias S, Nayak USL, Isaacs B. Balance in elderly patients: the "Get-up and Go Test". Arch Phys Med Rehabil 1986; 67: 387–389.
9. 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.
10. 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.
11. Olney SJ, Richards C. Hemiparetic gait following stroke. Part I: characteristics. Gait Posture 1996; 4: 136–148.
12. Roy G, Nadeau S, Gravel D, Piotte F, Malouin F, McFadyen BJ. Side difference in the hip and knee joint moments during sit-to-stand and stand-to-sit tasks in individuals with hemiparesis. Clin Biomech 2007; 22: 795–804.
13. Teixeira-Salmela L, Nadeau S, McBride I, Olney SJ. Effects of muscle strengthening and physical conditioning training on temporal, kinematic and kinetic variables during gait in chronic stroke survivors. J Rehabil Med 2001; 33: 53–60.
14. Malouin F. Observational gait analysis. In: Craik R, Oatis CA, editors. Gait analysis: theory and application. St Louis: Mosby; 1995, p. 112–124.
15. 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.
16. Toro B, Nester CJ, Farren PC. The status of gait assessment among physiotherapists in the United Kingdom. Arch Phys Med Rehabil 2003; 84: 1878–1884.
17. Toro B, Nester C, Farren P. A review of observational gait assessment in clinical practice. Physiother Theory Pract 2003; 19: 137–149.
18. Benson J, Clark F. A guide for instrument development and validation. Am J Occup Ther 1982; 36: 789–800.
19. Davis AE. Instrument development: getting started. J Neurosci Nurs 1996; 28: 204-207.
20. Streiner DL, Norman GR. Health measurement scales: a practical guide do their development and use. 3rd ed. New York: Oxford; 2004.
21. 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.
22. Bohannon RW. 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.
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. 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.
25. Centers for Disease Control and Prevention. Physical activity trends – United States, 1990–1998. Morb Mortal Wkly Rep 2001; 50: 166–169.
26. Grant JS, Davis LL. Selection and use of content experts for instrument development. Res Nurs Health 1997; 20: 269–274.
27. Berk RA. Importance of expert judgement in content-related validity evidence. West J Nurs Res 1990; 12: 659–671.
28. 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.
29. Sim J, Wright CC. The Kappa statistics in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005; 83: 257–268.
30. 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.
31. Toro B, Nester CJ, Farren PC. The development and validity of the Salford Gait Tool: An observaton-based clinical gait assessment tool. Arch Phys Med Rehabil 2007; 88: 321–327.
32. Toro B, Nester CJ, Farren PC. Inter- and intraobserver repeatability of the Salford Gait Tool: an observation-based clinical gait assessment tool. Arch Phys Med Rehabil 2007; 88: 328–332.
33. Janssen WGM, Bussmann HBJ, Stam HJ. Determinants of the sit-to-stand movement: a review. Phys Ther 2002; 82: 866–879.
34. Faria CDCM, Teixeira-Salmela L, Gomes Neto M, Rodrigues-de-Paula F. Performance-based tests in subjects with stroke: outcome scores, reliability and measurement errors. Clin Rehabil 2011; 26: 460–469.
35. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. New Jersey: Prentice-Hall; 2009.
36. Jakobsson U, Westergren A. Statistical methods for assessing agreement for ordinal data. Scand J Caring Sci 2005; 19: 427–431.
37. 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.
38. 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.
39. Capozzo A, Croce UD, Leardini A, Chiari L. Human movement analysis using stereophotogrammetry. Part 1: theoretical background. Gait Posture 2005; 21: 186–196.
40. States RA, Pappas E. Precision and repeatability of the Optotrack 3020 motion measurement system. J Med Eng Technol 2006; 30: 11–16.
41. Winter DA. Biomechanics and motor control of human movement. New York: John Wiley & Sons; 1990.
42. Gilleard W, Crobie J, Smith R. Rising to stand from a chair: symmetry, and frontal and transverse plane kinematics and kinetics. Gait Posture 2008; 27: 8–15.
43. Krebs DE, Wong D, Jevsevar D, Riley PO, Hodge WA. Trunk kinematics during locomotor activities. Phys Ther 1992; 72: 505–514.
44. Fortin C, Nadeau S, Labelle H. Inter-trial and test-retest reliability of kinematic and kinetic gait parameters among subjects with adolescent idiopathic scoliosis. Eur Spine J 2008; 17: 204–216.
45. Ghoussayni S, Stevens C, Durham S, Ewins D. Assessment and validation of a simple automated method for the detection of gait events and intervals. Gait Posture 2004; 20: 266–272.
46. Stokes VP, Anderson C, Forssberg H. Rotational and translational movement features of the pelvis and thorax during adult human locomotion. J Biomech 1989; 22: 43–50.
47. Dite W, Temple VA. Development of a clinical measure of turning for older adults. Am J Phys Med Rehabil 2002; 81: 857–866.
48. Thigpen M, Light KE, Creel GL, Flynn SM. Turning difficulty characteristics of adults aged 65 years or older. Phys Ther 2000; 80: 1174–1187.
49. Nadeau S, Desjardins P, Brière A, Roy G, Gravel D. A chair with a platform setup to measure the forces under each thigh when sitting, rising from a chair and sitting downs. Med Biol Eng Comput 2008; 46: 299–306.
50. Blake R, Lee S. The role of temporal structure in human vision. Behav Cogn Neurosci Rev 2005; 4: 21–42.
51. Gage JR, Ounpuu S. Gait analysis in clinical practice. Semin Orthop 1989; 4: 72–87.
52. Lynn MR. Determination and quantification of content validity. Nurs Res 1986; 35: 382–385.
53. 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.
54. Keenan AM, Bach TM. Video assessment of rearfoot movements during walking: a reliability study. Arch Phys Med Rehabil 1996; 77: 651–655.
55. Sim J, Arnell P. Measurement validity in physical therapy. Phys Ther 1993; 73: 102–115.

APPENDIX I. Final version of the Timed "Up and Go" Assessment of Biomechanical Strategies tool
Name:_____Date:_____

Chair specifications:_____Orthoses / Walking Aids: _____

SIT-TO-STAND		
A. Supporting the upper limb(s) with lateral trunk flexion and/or trunk rotation:		
( ) without support OR with support and no/slight trunk movement	( ) with support and moderate trunk movement	( ) with support and excessive trunk movement
B. Number of attempts to perform sit-to-stand and bringing the pelvis to the edge of the seat:		
( ) 1 without pelvis to edge of seat	( ) > 1 without pelvis to edge of seat	( ) > 1 with pelvis to edge of seat
C. Momentum generated by the first trunk flexion and extension of the trunk and lower limbs:		
( ) enough momentum for thighs-off and movements are continuous	( ) enough momentum for thighs-off and movements are NOT continuous	( ) not enough momentum for thighs-off
GAIT		
A. Step symmetry and length (most steps):		
( ) symmetrical and adequate length	( ) asymmetrical and adequate length on one side	( ) asymmetrical OR symmetrical and inadequate length on both sides
B. Initial contact with the heel (most steps)**:**		
( ) on both feet	( ) on one foot	( ) on neither feet
C. Hip extension during stance so that the thigh is posterior to the pelvis (most steps):		
( ) with both legs	( ) with one leg	( ) with neither leg
D. Foot clearance during swing phase (most steps):		
( ) with both feet	( ) with one foot	( ) with neither foot
E. Forward progression of the leg with atypical trunk movements (most steps)**:**		
( ) with both legs and no atypical trunk movements	( ) with one leg and atypical trunk movements	( ) with both legs and atypical trunk movements
TURN		
A. Step length between the outer and inner foot during the turn:		
( ) whole outer foot is ahead of inner foot	( ) part of outer foot is ahead of inner foot	( ) whole outer foot is beside/behind inner foot
B. Number of steps during the turn:		
( ) < 4	( ) 4-5	( ) > 5
C. Body rotations to accomplish the opposite direction during the turn:		
( ) < 3	( ) 3	( ) > 3
D. Sequencing of gait, turning, and gait:		
( ) movements are continuous and no loss of balance	( ) movements are NOT continuous and no loss of balance	( ) movements are NOT continuous and some loss of balance
STAND-TO-SIT		
A. Sequencing of gait, turning to sit, and stand-to-sit:		
( ) movements are continuous	( ) movements are mostly continuous	( ) movements are NOT continuous
B. Sequencing and control of thighs down and trunk back:		
( ) movements are continuous and good control of movements	( ) movements are NOT continuous and good control of movements	( ) movements are NOT continuous and poor control of movements
C. Parallel legs and knee flexion during stand-to-sit:		
( ) legs parallel and both knees flexion ≥ 90^0	( ) legs NOT parallel and both knees flexion ≥ 90^0	( ) knee flexion< 90^0 (one or both knees)
3 points for each category	2 points for each category	1 point for each category
Best performance		Worst performance
	Total score:_____/45	