

Development of the Modified Four Square Step Test and its reliability and validity in people with stroke

Margaret A. Roos, PT, PhD;^{1*} Darcy S. Reisman, PT, PhD;^{2–3} Gregory E. Hicks, PT, PhD;^{2–3} William Rose, PhD;² Katherine S. Rudolph, PT, PhD⁴

¹Department of Physical Therapy, University of the Sciences in Philadelphia, Philadelphia, PA; ²Graduate Program in Biomechanics and Movement Science and ³Department of Physical Therapy, University of Delaware, Newark, DE;

⁴College of Health Professions, University of New England, Portland, ME

Abstract—Adults with stroke have difficulty avoiding obstacles when walking, especially when a time constraint is imposed. The Four Square Step Test (FSST) evaluates dynamic balance by requiring individuals to step over canes in multiple directions while being timed, but many people with stroke are unable to complete it. The purposes of this study were to (1) modify the FSST by replacing the canes with tape so that more persons with stroke could successfully complete the test and (2) examine the reliability and validity of the modified version. Fifty-five subjects completed the Modified FSST (mFSST) by stepping over tape in all four directions while being timed. The mFSST resulted in significantly greater numbers of subjects completing the test than the FSST (39/55 [71%] and 33/55 [60%], respectively) ($p < 0.04$). The test-retest, intrarater, and interrater reliability of the mFSST were excellent (intraclass correlation coefficient ranges: 0.81–0.99). Construct and concurrent validity of the mFSST were also established. The minimal detectable change was 6.73 s. The mFSST, an ideal measure of dynamic balance, can identify progress in people with stroke in varied settings and can be completed by a wide range of people with stroke with the use of minimal equipment (tape, stop watch).

Key words: accidental falls, clinical testing, dynamic balance, gait speed, obstacle avoidance, outcome measure, reliability, static balance, stroke, validity.

INTRODUCTION

After stroke, as many as 50 percent of individuals fall while walking [1], and many of those individuals fall repeatedly [2]. Falling can result in serious consequences, including fractures and soft tissue injuries [3–5], increased fear of falling [6–9], and decreased activity [10], and can ultimately lead to placement in a long-term nursing facility [7]. Thus, a better understanding of the factors that contribute to falls after stroke might be possible by examining outcome measures that contain tasks that challenge dynamic balance.

Abbreviations: ABC = Activities-Specific Balance Confidence Scale, BBS = Berg Balance Scale, CI = confidence interval, FSST = Four Square Step Test, HHA = household ambulator, ICC = intraclass correlation coefficient, LCA = limited community ambulator, LI = less impaired, MDC = minimal detectable change, mFSST = Modified Four Square Step Test, SD = standard deviation, SI = significantly impaired, TUG = Timed “Up and Go” test, UCA = unlimited community ambulator.

*Address all correspondence to Margaret A. Roos, PT, PhD; University of the Sciences in Philadelphia, 600 South 43rd St, Philadelphia, PA 19104; 215-596-8676.

Email: m.roos@usciences.edu

<http://dx.doi.org/10.1682/JRRD.2014.04.0112>

Recently, Simpson et al. demonstrated that people with stroke are 1.8 times more likely to fall than age-matched control subjects and that they are more likely to fall at home [11]. While the Berg Balance Scale (BBS) was shown to be the strongest predictor of falls in control subjects, the strongest predictors of falls in people with stroke were worse BBS scores and faster Timed “Up and Go” test (TUG) scores. The authors noted that the TUG scores may have predicted falls in the stroke patients because faster TUG scores may represent a higher level of mobility that resulted in increased opportunities to fall [11]. This also illustrates that balance and mobility are important features leading to safe ambulation in people with stroke that warrant further investigation.

Obstacle avoidance, particularly when coupled with time constraints, increases the risk of falls in older adults [12]. An outcome measure of balance that requires obstacle avoidance under an imposed time constraint may be better able to identify dynamic balance deficits that contribute to an increased risk of falls in people with stroke. Such a test, the Four Square Step Test (FSST), was developed by Dite and Temple to examine dynamic balance in older adults [13]. The FSST is a timed measure that requires individuals to step over canes placed in a cross-wise pattern on the floor, thereby creating four quadrants as illustrated in **Figure 1(a)**. People are timed as they step clockwise, then counterclockwise, through each quadrant without touching the canes. Subjects are instructed to “face forward during the entire sequence, if possible” [13]. The FSST has been shown to be reliable and valid in community-dwelling older adults without disability and is fast and easy to administer [13]; thus, it is easily incorporated into clinical practice and research studies.

Interest in using the FSST to assess dynamic balance in people with stroke is growing. For example, Blennerhassett and Jayalath studied the FSST to assess changes in balance in ambulant people with stroke [14]. The authors reported that the FSST was a feasible and valid test for use in tracking the progress of persons during stroke rehabilitation [14]. However, when adopting a test to assess dynamic balance in people with hemiparesis following stroke, one must consider not only the validity and reliability of the test but also the number of individuals with stroke who can complete the test. This requires assessment of the test in people with stroke with a wide range of functional abilities. There is evidence that some people with stroke have difficulty completing the FSST. Blennerhassett and Jayalath reported that five subjects (5/37 [14%]) were unable to complete the

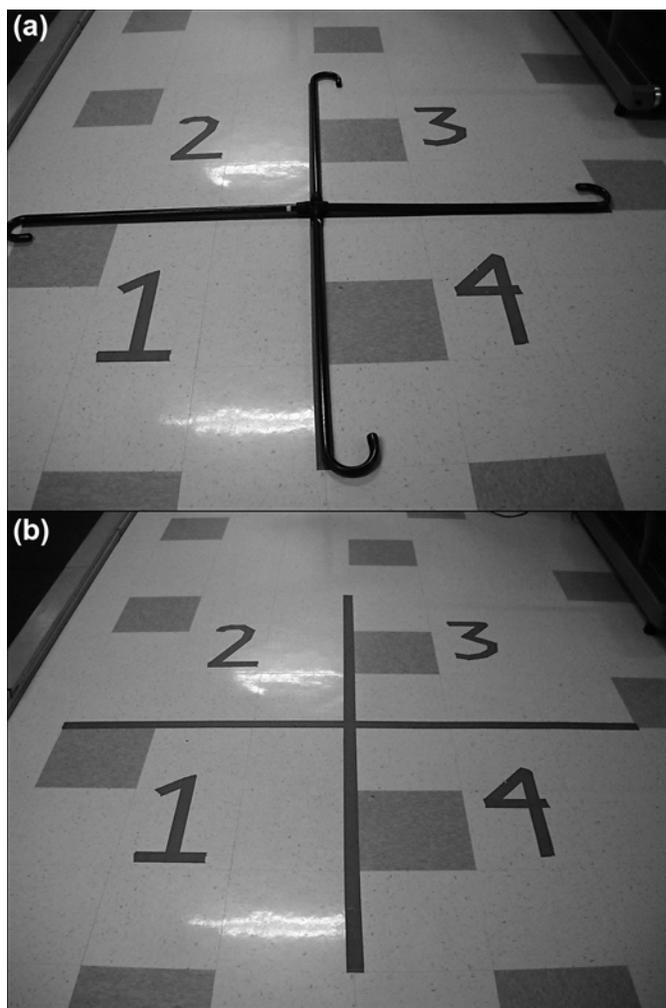


Figure 1. Setup of quadrants for (a) Four Square Step Test using canes and (b) Modified Four Square Step Test using tape.

test during the first testing session [14]. Moreover, subjects were given four attempts to complete the test, and 40 to 62 percent of the attempts were unsuccessful because of difficulty maintaining balance and clearing the canes. The authors also reported that some subjects had difficulty maintaining the face forward position when moving from one quadrant to the next. Goh et al. reported good intrarater and interrater reliability in a small sample of people with stroke, but they found that 20 percent of the subjects were unable to complete the test [15].

We undertook this study to modify the FSST to allow more people with stroke to complete the test while retaining the fundamental components of obstacle avoidance and the requirement that subjects step forward, backward,

and to each side. The modifications to the FSST included first asking subjects to face forward at all times during the test. Requiring subjects to face forward at all times ensures that an integral component of the test, stepping in multiple directions, is completed by all subjects. The second modification replaced the canes with tape on the ground, as shown in **Figure 1(b)**. This retained the need to step over an object but reduced the height of the object. We hypothesized that replacing the canes with tape would result in a valid and reliable test of dynamic balance that would allow significantly more people with stroke to complete the test.

METHODS

Community-dwelling people with stroke were recruited from local physical therapy clinics and stroke support groups and through advertisements in the local newspaper to participate in this cross-sectional study. Inclusion criteria included: age >18 yr, experienced one or more strokes, ability to walk without assistance of another person (orthotics and assistive devices were allowed), and ability to follow instructions and communicate with the investigators. Individuals were excluded if they had additional neurological diagnoses or had other significant health problems that affected walking ability. All subjects received medical clearance from their personal physician and signed informed consents that had been approved by the Human Subject Review Boards at the University of Delaware and Magee Rehabilitation Hospital in Philadelphia, Pennsylvania. The rights of all subjects were protected throughout the study.

Testing

Subjects attended up to two testing sessions in which they first completed the FSST by stepping over canes and then by stepping over 2 in. colored tape; we labeled the test using tape as the Modified Four Square Step Test (mFSST). During one of the sessions, the performance was videotaped. Subjects completed the FSST and mFSST by facing forward in quadrant 1, then stepping clockwise through quadrants 2 through 4, and then counterclockwise through quadrants 4 to 1 (**Figure 1**) with the instruction to “try to complete the sequence as fast as possible without touching the canes/tape. Both feet must make contact with the floor in each square. You need to face forward during the entire sequence. The clock will begin when I say ‘Step’ and stop when the last foot touches down in square 1 (finished). You

can take as many steps in each square as you need in order to make sure your feet clear the canes/tape.” Subjects were given one practice trial, and then the better score of two trials for both the FSST and mFSST was documented as their timed score. Subjects were unsuccessful in completing the trial if they made contact with the canes/tape, lost balance, or turned their body to face the next quadrant. The practice attempt allowed the tester to provide feedback to the subjects to ensure correct performance. Subjects used an assistive device and/or an orthotic if they used one to ambulate safely at home. To avoid fatigue, subjects took a standing or seated rest break between trials of the FSST and mFSST and were prompted to rest as often as they required.

Reliability

One rater (M.A.R.) established test-retest reliability from live performances during testing sessions that were separated by at least 1 d. Intrarater and interrater reliability were established using the videotaped performances evaluated by five raters (including M.A.R.) who were licensed physical therapists with between 8 mo and 24 yr (mean = 6.7 yr) of experience treating patients. Raters evaluated subjects’ performance by identifying whether subjects were successfully able to avoid stepping on canes/tape, turning their body, or losing balance (yes/no). If subjects were successful with the criteria, the raters provided a timed score.

To ensure a uniform skill level, raters attended a 15 min orientation in which the testing procedures and scoring criteria were reviewed. Rater competency was assessed using scores of videotaped performances of five subjects who were not part of the overall subject sample. Raters were deemed competent when identification of avoiding objects, turning, and loss of balance met the answers of the primary investigator (M.A.R.) and the timed score was within 2 s of M.A.R.’s score. Each rater rated the videotaped performances of each subject two times, separated by 7 d.

Validity

Validity is typically established by comparing scores on an outcome measure to a known criterion standard; however, no criterion standard exists for dynamic balance in people following stroke. When no criterion standard exists, concurrent validity and/or construct validity can be used to establish that the new tool is potentially more efficient, easier to administer, or more practical than a more established method [16]. Concurrent validity can be established when the scores of the measurement

tool and criterion tools measuring the same deficits are taken at the same time. Construct validity reflects the ability of an instrument to measure an abstract concept [16] and can be defined based on values from numerous outcome measures purported to evaluate the construct.

In this study, both concurrent and construct validity were determined. Concurrent validity was established by assessing the correlation between the mFSST and the Activities-Specific Balance Confidence Scale (ABC), BBS, and TUG, which were administered during the same testing session. Construct validity was established using methods similar to those of other investigators [17–18] that used multiple outcome measures to form the construct of interest. Specifically, the construct of significantly impaired (SI) dynamic balance was created using scores on the ABC, TUG, and BBS.

The ABC was important to include in this dynamic balance construct because it quantifies a person's opinion on the contribution of balance to performing functional tasks. The BBS was chosen because it measures static and dynamic balance and is commonly used in clinical and research settings. The TUG requires participants to rise, transport their body, turn while moving, and sit with adequate dynamic balance to receive a timed score. Including the TUG in the construct captures an objective measurement of challenging movements in the upright position.

Beninato et al. noted that an ABC score of 81.1 had a sensitivity of 1.00 and specificity of 0.72 (95% confidence interval [CI] = 0.51–0.93) to identify people with stroke who experience multiple falls [19]. Alzayer et al. reported that a BBS of 52 had a sensitivity of 0.90 (95% CI = 0.71–1.09) and specificity of 0.41 (95% CI = 0.24–0.58) to distinguish people with stroke who fell one time from those who fell multiple times [20]. Shumway-Cook et al. found that a TUG score over 13.5 s had a sensitivity of 0.80 and specificity of 1.00 to identify falls in older adults, although no such score has been established in people with stroke [21]. Many individuals with stroke have dynamic balance deficits. The cut-scores for the different outcome measures were chosen because they were able to differentiate between fallers in those with stroke (ABC, BBS) and older adults (TUG). Belgen et al. noted that people with stroke who were nonfallers and one-time fallers completed the TUG in 15.8 s on average, while multiple fallers required an average of 21.8 s [2]. However, Belgen et al. did not find significant differences between these two groups; therefore, the cut-score determined by Shumway-Cook et al. [21] was chosen because it was able to identify older adults who fell

and we expected a majority of our subjects to be older adults. Additionally, many people poststroke have impaired dynamic balance; therefore, the more stringent TUG cut-off score identified by Shumway-Cook et al. [21] would increase the likelihood of a worthwhile dynamic balance construct. Subjects who met all of the following criteria—ABC <81 percent, BBS <52, and TUG >13.5 s—were classified as having SI dynamic balance. Subjects who scored better than the cut-off on any one of the ABC, TUG, or BBS measures were classified as having less impaired (LI) dynamic balance. If statistically significant differences were observed between the SI and LI groups, the test was deemed valid for the FSST and mFSST.

Data were collected on the ability of subjects with a wide range of functional ability to receive a timed score on both the FSST and mFSST. To determine validity, we calculated the percentage of subjects who could perform each test and grouped them according to Perry et al.'s classification of function based on gait speed [22]. The classifications included household ambulators (HHAs) (gait speed <0.4 m/s) limited community ambulators (LCAs) (gait speed 0.4–0.6 m/s), and unlimited community ambulators (UCAs) (gait speed >0.8 m/s).

Minimal Detectable Change

Once the mFSST was determined to be valid and reliable, additional subjects were recruited to undergo testing of the mFSST on two separate occasions, and their results as well as the results from the subjects completing test-retest reliability were used to determine minimal detectable change (MDC). The formula used to calculate MDC with a 95 percent CI was based on the standard error of the mean [23]: $MDC_{95} = 1.96 \times SD_{baseline} \times \sqrt{2[1 - r_{test-retest}]}$ where SD was the standard deviation of the baseline measure and r was the Pearson product moment correlation.

Data Analysis

After determining the percentage of subjects who could complete the FSST and mFSST, the binomial test [24] was used to examine the hypothesis that the proportion of subjects who were able to complete the mFSST was greater than those who could complete the FSST. A one-sided analysis was used considering we removed the height of the obstacle and expected the test to be easier to complete.

To assess reliability, a kappa statistic was calculated to assess intrarater and interrater agreement for the categorical variables of avoiding canes/tape, turning, or losing

balance. Values of 0.61 to 0.80 indicated a substantial strength of agreement, while 0.81 to 1.00 was considered almost perfect agreement [25]. Intraclass correlation coefficients (ICCs) model 3,1 was used to assess test-retest and intrarater reliability, while ICC(2,1) was used to calculate interrater reliability for the timed score [16]. To determine the concurrent validity or the degree of association between outcome measures of balance (ABC, TUG, and BBS) and FSST and mFSST, the Spearman rank correlation coefficient was calculated. Following Colton's guidelines, a correlation coefficient between 0.50 to 0.75 was considered moderate, with values above 0.75 considered as good to excellent [26]

Group differences in the timed score of the FSST and mFSST between the SI and LI groups were evaluated using the Mann-Whitney U test. Adequate construct validity was established if the timed score in the SI group was significantly longer than that of the LI group at a level of $p < 0.05$.

RESULTS

Over an 18 mo period, 71 subjects were recruited to complete this study. Not all subjects completed each portion of the study. See **Table 1** for the subject number and demographics that participated in each portion. The principal investigator (M.A.R.) completed the test-retest reliability and MDC assessments. For the reliability assessments, videotapes of three subjects were excluded from the analy-

sis because of technical difficulties with the videotape, leaving videotapes from 26 subjects included in those reliability assessments. The numbers of subjects included to determine validity, test-retest reliability, and intra- and interrater reliability met or exceeded sample size estimates from a priori power analyses needed to detect statistically significant differences at $\beta = 0.80$ and $\alpha = 0.05$. Subjects had a wide range of mobility levels and utilized a variety of assistive devices and orthoses, which are described in **Table 1**. The range of functional abilities of the subjects can be seen in **Table 2**.

Of the 55 subjects, 33 (60%) were able to successfully complete the FSST, while 39 of 55 (71%) were able to complete the mFSST. Seven subjects who failed to complete the FSST were able to complete the mFSST, and one subject who completed the FSST could not complete the mFSST. The proportion of subjects who were able to complete the mFSST was significantly greater than those able to complete the FSST (binomial $p = 0.04$). Subjects failed to complete the FSST because of contact with the canes ($n = 21$), with one subject both contacting the canes and turning to face the next quadrant. Similarly, subjects who failed to complete the mFSST made contact with the tape ($n = 14$), with one subject making contact with the tape and turning to face the next quadrant and one subject making contact with tape and losing balance. The average times \pm SD required for subjects to complete the FSST and mFSST were 21.30 ± 14.57 s and 18.45 ± 12.30 s, respectively.

Table 1.
Demographics of subjects involved in different assessments.

Assessment	Age (yr), Mean (range)	Sex (n)	Time Since Stroke (mo), Mean (range)	Paretic Side (n)	Orthosis or Assistive Devices (n)
Validity ($n = 55$)	63.3 (39.8–87.4)	Male: 30 Female: 25	34.3 (1.8–136.5)	Right: 24 Left: 31	AFO: 7 SC: 10 QC: 4
Test-Retest Reliability ($n = 17$)	62.5 (47.6–80.0)	Male: 13 Female: 4	35.5 (5.6–124.9)	Right: 10 Left: 7	AFO: 3 SC: 1 QC: 3
Intra- and Interrater Reliability ($n = 26$)	63.3 (45.2–81.2)	Male: 17 Female: 9	35.3 (1.9–124.9)	Right: 11 Left: 15	AFO: 2 SC: 5 QC: 3
Minimal Detectable Change ($n = 36$)	60.0 (27.7–80.0)	Male: 24 Female: 12	29.9 (6.0–124.9)	Right: 14 Left: 22	AFO: 4 SC: 5 QC: 5

AFO = ankle foot orthosis, QC = quad cane, SC = straight cane.

Table 2.

Functional scores for subjects involved in different assessments. Numbers given as mean (range).

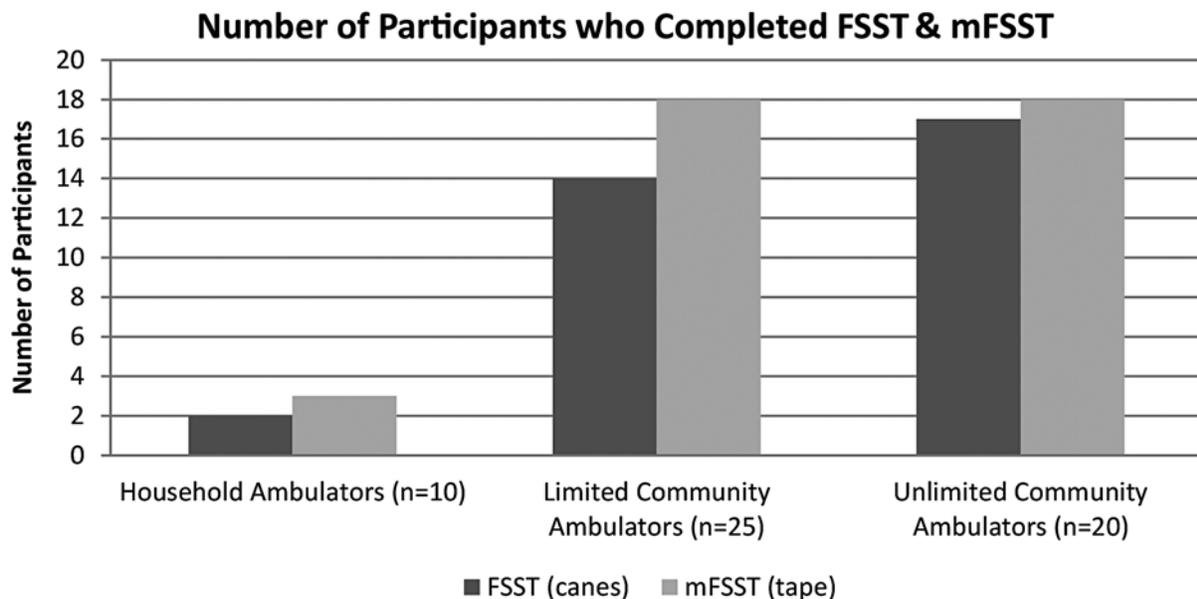
Assessment	Gait Speed (m/s)	BBS	ABC	TUG (s)	FSST (s)	mFSST (s)
Validity (<i>n</i> = 55)	0.70 (0.22–1.35)	50.4 (31.0–56.0)	76.6 (25.6–96.9)	14.72 (8.30–48.00)	21.30 (8.87–65.22)	18.45 (7.87–65.43)
Test-Retest Reliability (<i>n</i> = 17)	0.67 (0.22–1.07)	46.0 (29.0–55.0)	71.6 (48.8–95.6)	19.21 (9.13–54.70)	—	—
Intra- and Interrater Reliability (<i>n</i> = 26)	0.71 (0.22–1.07)	46.6 (29.0–56.0)	69.1 (20.0–95.6)	19.50 (9.13–54.70)	—	—

ABC = Activities-Specific Balance Confidence Scale, BBS = Berg Balance Scale, FSST = Four Square Step Test, mFSST = Modified Four Square Step Test, TUG = Timed “Up and Go” test.

When the 55 subjects who participated in this study were grouped according to Perry et al.’s classification of function based on gait speed [22], 10 subjects were classified as HHAs, 25 were classified as LCAs, and 20 were classified as UCAs. HHAs had the greatest difficulty with both forms of the FSST, with only 2/10 (20%) and 3/10 (30%) of the subjects being able to complete the FSST and mFSST, respectively. In the LCA group, 14/25 (56%) and 18/25 (72%) of subjects were able to complete the FSST and mFSST, respectively. In the UCA subjects, 17/20 (85%) and 18/20 (90%) were able to complete the FSST and mFSST, respectively (**Figure 2**).

Reliability

Intrarater and interrater reliability of the FSST and mFSST for the categorical variable of avoiding the obstacles ranged from substantial to almost perfect [25]. In the FSST and mFSST, a kappa coefficient could not be calculated for the subjects’ ability to avoid turning or loss of balance. In the FSST, no raters identified turning or loss of balance, while in the mFSST, no raters identified turning and only one rater for one attempt identified loss of balance. Test-retest reliability, intrarater reliability, and interrater reliability of the FSST and mFSST timed scores were excellent. See **Tables 3** and **4** for reliability scores.

**Figure 2.**

Number of subjects who completed Four Square Step Test (FSST) and Modified Four Square Step Test (mFSST).

Table 3.

Reliability of Four Square Step Test (FSST) and Modified Four Square Step Test (mFSST) for avoiding obstacles.

Reliability*	Kappa Statistic (95% CI)	
	FSST	mFSST
Intrarater		
Range Bottom	0.91 ± 0.06 (0.84–0.97)	0.92 ± 0.06 (0.84–0.97)
Range Top	1.0 ± 0.0 (1.0–1.0)	1.0 ± 0.0 (1.0–1.0)
Interrater		
Range Bottom	0.74 ± 0.10 (0.64–0.84)	0.91 ± 0.06 (0.85–0.97)
Range Top	1.0 ± 0.0 (1.0–1.0)	1.0 ± 0.0 (1.0–1.0)

*For raters 1–5.
CI = confidence interval.

Validity

The FSST and mFSST were highly correlated with each other (Spearman $\rho = 0.954$, $p < 0.001$). Concurrent validity was established for the FSST because there were moderate negative correlations between the time to complete the FSST and the BBS (Spearman $\rho = -0.685$, $p < 0.001$) and the ABC (Spearman $\rho = -0.538$, $p < 0.001$) and a strong positive relationship between the time to complete the FSST and the TUG (Spearman $\rho = 0.727$, $p < 0.001$). Concurrent validity of the mFSST was established by moderate negative correlations between the time to complete the mFSST and the BBS (Spearman $\rho = -0.616$, $p < 0.001$) and the ABC (Spearman $\rho = -0.449$, $p = 0.004$), as well as a strong positive correlation between the time to complete the mFSST and TUG times (Spearman $\rho = 0.726$, $p < 0.001$).

Considering the FSST and using the impaired balance construct created by ABC, BBS, and TUG criteria, 8 subjects were classified as SI, with a score of 37.5 ± 18.5 s, while 25 subjects were classified as LI, with a score of 16.1 ± 8.3 s. Of those that completed the mFSST, 10 subjects were classified as SI, with a score of 30.8 ± 17.2 s, while 29 were classified as LI, with a timed score of 14.2 ± 6.1 s. The construct validity measures of FSST and mFSST were established by the fact that the times to complete the tests were significantly longer for the SI subjects than the LI subjects (both $p < 0.001$).

Minimal Detectable Change

Of the 36 subjects, 28 were able to complete the mFSST at both testing times and were included in the MDC analysis. The MDC was calculated to be 6.73 s.

Table 4.

Reliability of Four Square Step Test (FSST) and Modified Four Square Step Test (mFSST) for timed scores.

Reliability	ICC _{3,1} (95% CI)	
	FSST	mFSST
Test-Retest*	0.85 (0.27–0.99)	0.90 (0.68–0.97)
Intrarater†	0.99 (0.99–0.99)	>0.81 (0.81–0.99)
Interrater†	0.99 (0.99–1.00)	>0.81 (0.68–0.97)

*For rater M.A.R.
†For raters 1–5.
CI = confidence interval, ICC = intraclass correlation coefficient.

DISCUSSION

Our results demonstrate that the mFSST, which was able to be completed by a significantly larger number of persons with stroke than the FSST, has excellent test-retest, intrarater, and interrater reliability as well as construct and concurrent validity. Our observation that individuals with stroke had difficulty completing the FSST was consistent with findings of Blennerhassett and Jayalath, who found that 62 percent of patients had difficulty clearing the canes at least one time over the three time points outlined in their study [14]. They also reported that 5 of the 37 subjects (14%) were unable to complete the test at the start of the study, which is a substantially smaller proportion of subjects than those who were unable to perform the test in the present study. This is likely due to the lower functional level of the subjects in our study. Compared with previous studies of the FSST in persons with stroke, our subjects were more impaired, with an average time of 37.5 s to complete the test. This is compared with the subjects in Blennerhassett and Jayalath [14], whose timed scores averaged between 17 and 21 s and Goh et al. [15], whose subjects completed the FSST in an average of 17.4 s.

This information reinforces the idea that an outcome measure is necessary to capture the difficulty with obstacle avoidance in individuals who are more functionally impaired. The mFSST is such a test since it is able to reliably capture the difficulty experienced by people with stroke when asked to step over obstacles in multiple directions. As a result of our investigation, we also demonstrated that the FSST with the modified instructions to face forward at all times was a reliable and valid measure of dynamic balance in people with stroke.

When evaluating the concurrent validity between the FSST and BBS, our findings are different than those

noted by Goh et al., who found no correlation between these two outcome measures [15]. Potential reasons for this difference could be that Goh et al. studied 15 people with stroke who used a cane to complete the measure [15], while only 14 of the 55 subjects in this study used a cane for mobility, implying that more of our subjects had greater functional mobility. Additionally, the difference in subject numbers may have resulted in subjects with a greater range of mobility, which captured the significantly negative relationship between the BBS and FSST.

The influence of functional level on the number of subjects able to complete the FSST and mFSST is also supported by the fact that only 20 and 30 percent, respectively, of the subjects in the HHA category could complete the tests. This large floor effect for people poststroke with more limited functional ability suggests that a test of dynamic balance that involves single-limb stance and moving the opposite limb forward, backward, or sideways may be too challenging. The effect of modifying the FSST by asking individuals to clear a taped line is most notable in the percentage of subjects in the LCA category who could now complete the test. The fact that obstacle avoidance is very difficult for people in this category and that this can be captured by this simple modification is advantageous to clinicians and researchers. Nearly all subjects in the UCA group, whose functional mobility is highest, were able to complete both forms of the FSST, as might be expected. For individuals in the LCA and UCA categories, the timed score will be able to capture changes in a wide range of deficits throughout the rehabilitation process.

In our testing of the FSST and the mFSST, the instructions given included the instruction to maintain balance and complete the test as quickly as possible while facing forward and avoiding contact with the canes/tape. The primary reason that subjects failed to complete the test was contact with obstacles. Very few subjects failed due to turning their body or loss of balance, which may be due to the demonstration, instructions, and reinforcements given during testing. When the subjects made contact with the objects on the floor, the therapist stopped timing the test and when necessary assisted the subjects to return to the starting position or to a seated position. This behavior by the therapist may have resulted in the lack of identification of failure to complete the test because of loss of balance because contact with the objects came first. The results of our investigation demonstrate that both FSST and mFSST with

these instructions are reliable and valid measures of dynamic balance in people poststroke.

The mFSST is an outcome measure that is easy to administer, requires minimal space and equipment (tape and a stop watch), and can be completed in a short period of time. These are all very desirable features of an outcome measure. When comparing the mFSST with commonly used balance outcome measures such as the BBS and the TUG in people with stroke, the mFSST has some advantages. The BBS requires considerably longer to complete and the tasks included may not be challenging enough to detect dynamic balance deficits in those who can ambulate that can be detected with the mFSST. The TUG requires a similar time commitment, but the advantage of the mFSST is the increased level of difficulty because it includes obstacle avoidance and a time constraint, which are both difficult conditions for people with stroke.

This study has limitations. One limitation is that the order of completion of the FSST and the mFSST was not randomized. The FSST, being the more difficult test, was completed first to decrease the likelihood that subjects would have developed a plan to lift their lower limb a short distance to clear the tape, which would be insufficient to clear the canes, resulting in an immediate “not able to be completed” score on the first attempt at the FSST. Another limitation is the number of subjects studied with the most limited gait speeds. It would be beneficial to study individuals in this gait speed category more extensively to find whether greater numbers would alter the findings and/or find a dynamic balance measure that requires obstacle avoidance that can be completed by these people with stroke.

CONCLUSIONS

The mFSST is a timed measure in which individuals step over tape lines in different directions. The mFSST demonstrated strong test-retest, intrarater, and interrater reliability as well as construct and concurrent validity in people with stroke with a wide range of gait speeds. Early identification of an individual’s dynamic balance difficulties during walking is the first step in assisting clinicians in choosing appropriate activities during rehabilitation to decrease the high incidence of falls in people with stroke early after discharge to home. The mFSST can be completed by a larger proportion of persons with

stroke than the FSST, particularly for individuals classified as LCAs. This makes the mFSST a good alternative in those poststroke with all but the most profound functional limitations.

ACKNOWLEDGMENTS

Author Contributions:

Conceived study: M. A. Roos, D. S. Reisman, K. S. Rudolph.

Collected data: M. A. Roos.

Processed data: M. A. Roos.

Analyzed data: M. A. Roos, D. S. Reisman, K. S. Rudolph, G. E. Hicks, W. Rose.

Prepared manuscript: M. A. Roos.

Edited manuscript: D. S. Reisman, K. S. Rudolph, G. E. Hicks, W. Rose.

Financial Disclosures: The authors have declared that no competing interests exist.

Funding/Support: This material was based on work supported by the National Institutes of Health (grant R010786).

Institutional Review: All subjects signed informed consent that had been approved by the Human Subject Review Boards at the University of Delaware and Magee Rehabilitation Hospital in Philadelphia, Pennsylvania. The rights of all subjects were protected throughout the study.

Participant Follow-Up: The authors plan to inform participants of the results of this study.

REFERENCES

- Harris JE, Eng JJ, Marigold DS, Tokuno CD, Louis CL. Relationship of balance and mobility to fall incidence in people with chronic stroke. *Phys Ther*. 2005;85(2):150–58. [PMID:15679466] <http://dx.doi.org/10.1161/hs0202.102375>
- Belgen B, Beninato M, Sullivan PE, Narielwalla K. The association of balance capacity and falls self-efficacy with history of falling in community-dwelling people with chronic stroke. *Arch Phys Med Rehabil*. 2006;87(4):554–61. [PMID:16571397] <http://dx.doi.org/10.1016/j.apmr.2005.12.027>
- Kerse N, Parag V, Feigin VL, McNaughton H, Hackett ML, Bennett DA, Anderson CS; Auckland Regional Community Stroke (ARCOS) Study Group. Falls after stroke: Results from the Auckland Regional Community Stroke (ARCOS) Study, 2002 to 2003. *Stroke*. 2008;39(6):1890–93. [PMID:18483413] <http://dx.doi.org/10.1161/STROKEAHA.107.509885>
- Jørgensen L, Engstad T, Jacobsen BK. Higher incidence of falls in long-term stroke survivors than in population controls: Depressive symptoms predict falls after stroke. *Stroke*. 2002;33(2):542–47. [PMID:11823667] <http://dx.doi.org/10.1161/hs0202.102375>
- Stel VS, Smit JH, Pluijms SM, Lips P. Consequences of falling in older men and women and risk factors for health service use and functional decline. *Age Ageing*. 2004;33(1):58–65. [PMID:14695865] <http://dx.doi.org/10.1093/ageing/afh028>
- Schmid AA, Rittman M. Consequences of poststroke falls: Activity limitation, increased dependence, and the development of fear of falling. *Am J Occup Ther*. 2009;63(3):310–16. [PMID:19522139] <http://dx.doi.org/10.5014/ajot.63.3.310>
- Yates JS, Lai SM, Duncan PW, Studenski S. Falls in community-dwelling stroke survivors: An accumulated impairments model. *J Rehabil Res Dev*. 2002;39(3):385–94. [PMID:12173758]
- O’Loughlin JL, Robitaille Y, Boivin JF, Suissa S. Incidence of and risk factors for falls and injurious falls among the community-dwelling elderly. *Am J Epidemiol*. 1993;137(3):342–54. [PMID:8452142]
- Lord SR, Ward JA, Williams P, Anstey KJ. An epidemiological study of falls in older community-dwelling women: The Randwick falls and fractures study. *Aust J Public Health*. 1993;17(3):240–45. [PMID:8286498] <http://dx.doi.org/10.1111/j.1753-6405.1993.tb00143.x>
- Mackintosh SF, Hill K, Dodd KJ, Goldie P, Culham E. Falls and injury prevention should be part of every stroke rehabilitation plan. *Clin Rehabil*. 2005;19(4):441–51. [PMID:15929514] <http://dx.doi.org/10.1191/0269215505cr796oa>
- Simpson LA, Miller WC, Eng JJ. Effect of stroke on fall rate, location and predictors: A prospective comparison of older adults with and without stroke. *PLoS ONE*. 2011;6(4):e19431. [PMID:21559367] <http://dx.doi.org/10.1371/journal.pone.0019431>
- Weerdesteyn V, Nienhuis B, Geurts AC, Duysens J. Age-related deficits in early response characteristics of obstacle avoidance under time pressure. *J Gerontol A Biol Sci Med Sci*. 2007;62(9):1042–47. [PMID:17895445] <http://dx.doi.org/10.1093/gerona/62.9.1042>
- Dite W, Temple VA. A clinical test of stepping and change of direction to identify multiple falling older adults. *Arch Phys Med Rehabil*. 2002;83(11):1566–71. [PMID:12422327] <http://dx.doi.org/10.1053/apmr.2002.35469>
- Blennerhassett JM, Jayalath VM. The Four Square Step Test is a feasible and valid clinical test of dynamic standing balance for use in ambulant people poststroke. *Arch Phys Med Rehabil*. 2008;89(11):2156–61. [PMID:18996245]
- Goh EY, Chua SY, Hong SJ, Ng SS. Reliability and concurrent validity of Four Square Step Test scores in subjects with chronic stroke: A pilot study. *Arch Phys Med Rehabil*. 2002;33(2):542–47. [PMID:11823667] <http://dx.doi.org/10.1161/hs0202.102375>

- 2013;94(7):1306–11. [PMID:23416218]
<http://dx.doi.org/10.1016/j.apmr.2013.01.027>
16. Portney LG, Watkins MP. Foundations of clinical research: Applications to practice. 3rd ed. Upper Saddle River (NJ): Prentice Hall Health; 2009.
17. Hicks GE, Manal TJ. Psychometric properties of commonly used low back disability questionnaires: Are they useful for older adults with low back pain? *Pain Med.* 2009;10(1):85–94. [PMID:19222773]
<http://dx.doi.org/10.1111/j.1526-4637.2008.00548.x>
18. Smith BH, Penny KI, Purves AM, Munro C, Wilson B, Grimshaw J, Chambers WA, Smith WC. The Chronic Pain Grade questionnaire: Validation and reliability in postal research. *Pain.* 1997;71(2):141–47. [PMID:9211475]
[http://dx.doi.org/10.1016/S0304-3959\(97\)03347-2](http://dx.doi.org/10.1016/S0304-3959(97)03347-2)
19. Beninato M, Portney LG, Sullivan PE. Using the International Classification of Functioning, Disability and Health as a framework to examine the association between falls and clinical assessment tools in people with stroke. *Phys Ther.* 2009;89(8):816–25. [PMID:19520733]
<http://dx.doi.org/10.2522/ptj.20080160>
20. Alzayer L, Beninato M, Portney LG. The accuracy of individual Berg Balance Scale items compared with the total Berg score for classifying people with chronic stroke according to fall history. *J Neurol Phys Ther.* 2009;33(3):136–43. [PMID:19809392]
<http://dx.doi.org/10.1097/NPT.0b013e3181b51307>
21. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. *Phys Ther.* 2000;80(9):896–903. [PMID:10960937]
22. Perry J, Garrett M, Gronley JK, Mulroy SJ. Classification of walking handicap in the stroke population. *Stroke.* 1995;26(6):982–89. [PMID:7762050]
<http://dx.doi.org/10.1161/01.STR.26.6.982>
23. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther.* 2006;86(5):735–43. [PMID:16649896]
24. Zar JH. Biostatistical analysis. 5th ed. Upper Saddle River (NJ): Prentice Hall; 2010.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–74. [PMID:8435711]
<http://dx.doi.org/10.2307/2529310>
26. Colton T. Statistics in medicine. Boston (MA): Little Brown; 1974.

Submitted for publication April 24, 2014. Accepted in revised form September 23, 2015.

This article and any supplementary material should be cited as follows:

Roos MA, Reisman DS, Hicks GE, Rose W, Rudolph KS. Development of the Modified Four Square Step Test and its reliability and validity in people with stroke. *J Rehabil Res Dev.* 2016;53(3):403–12.

<http://dx.doi.org/10.1682/JRRD.2014.04.0112>

