Gait & Posture 72 (2019) 96–101

Contents lists available at ScienceDirect

Gait & Posture

journal homepage: www.elsevier.com/locate/gaitpost

Test-retest reliability and minimal detectable change for measures of balance and gait in adults with cerebral palsy

Ilana Levin[a,*], Michael D. Lewek[b], Carol Giuliani[b], Richard Faldowski[c], Deborah E. Thorpe[b]

[a] Human Movement Science Curriculum, Department of Allied Health Sciences, The University of North Carolina at Chapel Hill, NC, USA
[b] Division of Physical Therapy, Department of Allied Health Sciences, The University of North Carolina at Chapel Hill, NC, USA
[c] Department of Allied Health Sciences, The University of North Carolina at Chapel Hill, NC, USA

ARTICLE INFO

Keywords:
Cerebral palsy
Balance outcome measures
Gait variables
Test-retest reliability
Minimal detectable change

ABSTRACT

Background: Walking and balance often begin to deteriorate in ambulant adults with cerebral palsy (CP) in early adulthood. The decline in walking and balance imposes a more sedentary lifestyle, increases falls risk, negatively affects health, participation, and quality of life, and ultimately results in increased disability. Available research is not sufficient to guide interventions to improve walking and balance in this population. To advance research in this area, there is a need for measures of gait and balance with proven psychometrics for adults with CP.
Research question: The goal of this study was to determine test-retest reliability and minimal detectable change (MDC) values and to assess score distribution for the Balance Evaluation Systems Test (BESTest) and the Four Square Step Test (FSST) as measures of balance, for the Activities-specific Balance Confidence (ABC) Scale and the Modified Fall Efficacy Scale (MFES) as measures of balance confidence, and for over-ground spatiotemporal gait parameters at comfortable gait speed (CGS) and fast gait speed (FGS).
Methods: Twenty ambulant adults with CP (mean age 32.7 years), GMFCS-E&R Levels I and II, were tested twice within an average of 10 days. Test–retest reliability was evaluated using intra-class correlation coefficients ($ICC_{2,1}$), and MDC_{95} values were calculated using standard error of measurement values.
Results: The test-retest reliability of most outcome measures was good to excellent. ICC values were: BESTest = 0.99, BESTest sections 0.88 to 0.98, FSST = 0.91, ABC = 0.86, MFES = 0.9, CGS = 0.88, and FGS = 0.98. MDC values were: BESTest total = 4.9%, BESTest sections 8.7%–21.2%, FSST = 3.7 s, ABC = 18%, MFES = 1 point, CGS = 0.26 m/s, and FGS = 0.14 m/s. Most outcome scores were broadly distributed over scales ranges.
Significance: Adults with CP demonstrated stable test-retest performance on the selected measures. These measures could be useful to assess balance and gait of adults with CP. The MDC values can help evaluate whether observed changes exceed the expected random test-retest variations.

1. Introduction

An anoxic brain injury occurring prior, during, or shortly after birth leads to the diagnosis of cerebral palsy (CP) which is defined as a group of disorders of the development of movement and posture which cause activity limitations throughout life [1]. Approximately 60% of school-age children with CP are able to walk independently, but exhibit balance deficits that continue to affect their function into adulthood [2]. Life expectancy for most adults with CP is nearing that of the general population [3]. Although not a progressive neurological disorder, more than 50% of ambulant adults with CP experience a decline in balance and walking abilities in their twenties or thirties [2]. This decline can lead to a greater falls risk, a more sedentary lifestyle, and increase disability [2,4,5].

There is a paucity of intervention studies to improve dynamic balance and walking in adults with CP [6,7]. To design intervention protocols and interpret outcomes, we require standardized outcome measures that demonstrate, in adults with CP, good test-retest reliability, which confirms score consistency in the absence of intervention. We also require to know the minimal detectable change (MDC) values that inform about the magnitude of score change that can be attributed to a true change in capacity, that is beyond random performance variations [8]. Previous studies in ambulant individuals with CP used the Balance Evaluation Systems Test (BESTest), which includes the common "stand-

* Corresponding author.
E-mail addresses: ilevin@med.unc.edu (I. Levin), mlewek@med.unc.edu (M.D. Lewek), carol_giuliani@med.unc.edu (C. Giuliani), rafaldow@med.unc.edu (R. Faldowski), dthorpe@med.unc.edu (D.E. Thorpe).

https://doi.org/10.1016/j.gaitpost.2019.05.028
Received 19 December 2018; Received in revised form 14 May 2019; Accepted 26 May 2019
0966-6362/

Table 1
Demographic and functional characteristics of participants (n = 20).

	All participants			GMFCS-E&R I		GMFCS-E&R II	
	Mean (SD)		Range				
Gender (male/ female)				2/3		3/12	
Age (years)	32.7	(9.3)	21-50	31.8	(11.5)	32.9	(8.8)
Height (cm)	166	(11.1)	145-196	173	(4.3)	163.8	(11.8)
Weight (kg)	72.3	(16.8)	53-111	78	(24.6)	70.5	(13.9)
Topographical classifications:							
Spastic hemiplegia				3		2	
Spastic diplegia				2		12	
Spastic quadriplegia				–		1	
Assistive devices:							
Crutches/cane				–		2/2	
Shoe inserts and/or braces				1		3	

alone" tests of dynamic balance - the Functional Reach Test (FRT) and Timed Up and Go test (TUG) [9,10], the Activities-specific Balance Confidence scale (ABC), the Fall Efficacy Scale (FES) [11], spatiotemporal gait parameters [10], and the Four Square Step Test (FSST) [12]. However, the psychometric properties of these measures have not been established for adults with CP. The purpose of this study was to establish estimates of test-retest reliability and MDC values and to assess the distribution of scores over the ranges of the scales of measures of balance, balance confidence, falls efficacy, and spatiotemporal gait parameters in adults with CP. We hypothesized that test-retest reliability would be high and MDC values would be clinically acceptable, and that scores would be broadly spread along the scales of the measures.

2. Methods

2.1. Participants

Twenty participants (Table 1), completed the test-retest protocol ("Test 1", "Test 2") within an average of 10 days between tests (SD = 4.5, range: 6–23 days). Inclusion criteria were age 18–55 years, diagnosis of spastic CP, classified by a physical therapist as Gross Motor Function Classification Scale - Extended and Revised (GMFCS-E&R) [13] Level I (able to walk in all settings with some balance and coordination impairments) or Level II (walking is limited in some settings), able to walk for 10 min on level ground with no more than one rest break, and able to understand and follow simple instructions in English. Exclusion criteria included conditions that limit the ability to exercise, like recurrent episodes of knee or hip pain in the previous three months, uncorrected vision or hearing, vestibular disorders, uncontrolled seizures, second or third trimester pregnancy, or a significant restriction in communication. Participants were recruited through hospital clinics and research databases. This study was approved by the Committee for the Protection of Human Subjects at the University of North Carolina at Chapel Hill. All subjects provided informed consent prior to participation.

2.2. Outcome measures

All participants provided demographic and functional status information. All outcome measures were repeated in the same order in both tests. Participants that required forearm crutches or a cane for specific test items, used them in both tests. Scoring was done by a licensed physical therapist with 20 years of clinical experience (IL). In order to estimate population performance consistency on the selected measures between the two tests, while minimizing scoring errors, the scores collected during the tests were verified using videotapes. In rare cases of stopwatch failure, the participant was not asked to repeat the item, instead, videotaped-based times were used for analysis. To ensure assessment fidelity, the rater completed standardized online training using the BESTest website prior to the study [14]. Participants were assessed on the following outcome measures in order:

1 Gait parameters: Participants completed four passes along a 20 ft. Zeno Walkway System[1] at a comfortable gait speed (CGS) and 4 passes at a fast gait speed (FGS). For CGS, participants were instructed to "walk at your comfortable walking speed, as if going to get somewhere, but not rushing", and for FGS, they were instructed to "walk as fast as possible while maintaining safety". Data were processed using ProtoKinetics Movement Analysis Software[i] [15,16], and the following gait parameters were extracted: average cadence, stride length, longer and shorter step lengths, stride width, double support time, longer and shorter percentage swing times, and coefficients of variation for step and stride lengths, stride width, and gait phases percentage times (see supplementary material for definitions).
2 ABC scale: Participants completed the ABC scale independently, rating their confidence in performing 16 activities "without losing balance or becoming unsteady" on a 0% (no confidence) to 100% (complete confidence) 11-point scale. The responses were averaged to produce the ABC total score [17].
3 MFES: Participants completed the MFES independently, rating their confidence in performing 14 activities "without falling" on a 0 (no confidence) to 10 (complete confidence) 11-point scale. The responses were averaged to produce an MFES total score [18].
4 FSST: Following one to three untimed practice trials, participants were timed performing the multidirectional stepping task twice and the fastest time was used for analysis. Instructions were given and time was measured according to the published protocol [19].
5 BESTest: Participants completed 36 items, which were organized into 6 sections (Biomechanical Constraints, Stability Limits, Anticipatory Postural Adjustments, Postural Responses, Sensory Orientation, and Stability in Gait). Short rests were provided at the end of each section, and as needed. The BESTest was administered in accordance with published procedures [9] with a few modifications: (1) For all Sensory Orientation items, participants stood within their usual base of support width and not with feet together. Based on pilot data, testing with feet together would have resulted in many participants scoring 0 on most items, which would have reduced the clinical usefulness of this section. (2) Standing on the foam with eyes open and then eyes closed was done without stepping off, unless the participant required rest. (3) For all items, if participants did not initially follow instructions, secondary instructions were given and trials repeated. Each item was scored using a 4-point scale from 0 (worst performance) to 3 (best performance). Scores were summed and calculated as a percentage of the achievable score for each section and for the total score [9]. The best measured values of the FRT and the TUG were scored (0–3) as part of the BESTest. In addition, the collected values of distance and time were used for a separate analysis to obtain ICC estimates and MDC values for each of the tests.

2.3. Data analysis

Sample size: An a priori power analysis [20] indicated that a sample size of 19 was sufficient to establish that a detected reliability coefficient above 0.80 (good reliability) was significantly different from a reliability coefficient below 0.5 (cut off value for poor reliability) with a

[1] ProtoKinetics LLC, Havertown, PA

power of 0.80 and an alpha of 0.05 [21].

For all outcome measures, the following values were calculated: the intraclass correlation coefficient for test-retest absolute agreement for a single random rater ($ICC_{2,1}$) and the respective 95% confidence interval, the standard error of measurement (SEM)[2], and the derived MDC_{95}[3] [8]. Test-retest reliability was interpreted as “poor” if ICC was below 0.5, “moderate” if $ICC_{2,1}$ was 0.5 to 0.75, “good” if $ICC_{2,1}$ was 0.75 to 0.9, and excellent if $ICC_{2,1}$ was above 0.90 [22].

Test 1 and Test 2 scores were plotted relative to a unity line and visually inspected to assess range and proximity to the unity line. Systematic differences were identified using a paired *t*-test. For all data analyses, the alpha was set at 0.05. All statistical calculations were completed using R v.3.4.4 statistical software [23].

3. Results

3.1. Balance outcome measures

Tests means, $ICC_{2,1}$ and MDC_{95} values are presented in Table 2. Test-retest reliability values for the BESTest scores were “good” to “excellent” ($ICC_{2,1}$ values 0.88 to 0.99), “excellent” for the forward FRT ($ICC_{2,1}$ = 0.90), “good” for the FRT to both sides ($ICC_{2,1}$ = 0.78), “excellent” for the TUG ($ICC_{2,1}$ = 0.97), “excellent” for the FSST ($ICC_{2,1}$ = 0.91), “good” for the ABC ($ICC_{2,1}$ = 0.86), and “excellent” for the MFES ($ICC_{2,1}$ = 0.90).

The TUG scores (Test 1: 21 s, Test 2: 15.5 s) of one outlying participant reflected a pronounced learning effect related to confidence getting up from and down to the chair, and were omitted from analysis. Two participants performed the FSST exceptionally slowly (> 30 s) and were omitted from FSST analysis as outliers to prevent MDC inflation. Following the outlier omissions, all analyzed balance scores demonstrated non-significant paired t-tests, reflecting the absence of systematic difference between tests (Table 2).

The scores on most BESTest section score plots (Fig. 1a) were distributed broadly over the ranges of the scales and were close to the unity line. In section II (Stability Limits), all participants scored above 65%, and in section V (Sensory Orientation), 16/20 participants scored above 80%. On the TUG (Fig. 1b), and the FSST (Fig. 1d), the lower performing participants appeared to demonstrate slightly better scores on Test 2, but the differences between tests for the whole sample were non-significant (Table 2). Measured FRT distances captured broad performance ranges. Forward reach appeared to have a more consistent test-retest performance than reach to the sides (Fig. 1c) which was reflected in the relatively higher ICC values (Table 1). ABC scores were distributed broadly over the range of the scale, while MFES scores were clustered in higher range of the scale, with all participants scoring above 60% (Fig. 2).

3.2. Gait parameters

For each participant, an average of 36.2 (SD = 9.2) steps were analyzed for CGS, and 30.5 (SD = 8.8) steps for FGS. Tests means, $ICC_{2,1}$ and MDC_{95} values are presented in Table 3. Test-retest reliability was “good” ($ICC_{2,1}$ = 0.88) for CGS and “excellent” ($ICC_{2,1}$ = 0.98) for FGS. At both walking speeds, all spatiotemporal variables had “good” to “excellent” test-retest reliability. Step and stride length variables demonstrated small significant increase in values from Test 1 to Test 2 for both speeds. All other variables had non-significant paired t-tests, reflecting the absence of systematic difference between tests (Table 3). The coefficients of variation of the gait variables had mostly “moderate” ICC values, which appeared to be lower at FGS (see supplementary material for detailed results and discussion of gait variability metrics).

4. Discussion

This study provided evidence related to test-retest reliability and MDC_{95} values of balance and gait outcome measures for ambulatory adults with CP. Our results suggest that all balance measures and most spatiotemporal gait variables had “good” to “excellent” test-retest reliability, reflecting the tools’ ability to provide consistent test-retest measurements of performance.

Established MDC values can help identify a true change in measured performance that is beyond random variations [8]. As a derivative of the intraclass correlation and the standard deviation of the scores, the MDC value provides some insights into the psychometrics of the outcome measure. In this study, the MDC_{95} for the BESTest total score had a smaller value than the scores of its individual sections, reflecting an overall reliable tool that is stable and has little measurement error to conceal a true change in the measured capacity. Two sections (Reactive Balance and Sensory Orientation) had large MDC values, despite “excellent” and “good” reliability, reflecting the nonsystematic test-retest variations (Fig. 1a). High MDC values may raise concerns regarding the precision of the outcome measure. For example, high MDC values may indicate that either the assessed capacity is not stable from day to day, that the outcome measure itself is not specific enough to measure the true capacity, or that the measured performance is affected by other factors [24].

In a population that demonstrates a broad range of scores on a given outcome measure, the use of an MDC value may have some limitations when evaluating changes in individuals with either very high or very low scores. On the FSST, the fastest participant scored 6.8 s. For this participant, lowering FSST time may not be a clinical goal. There is no FSST cutoff score that indicates high falls risk in adults with CP to identify those who may benefit from lowering FSST time by at least the MDC_{95} value (3.7 s). Additionally, the FSST has a cognitive processing component that may affect performance with repeated exposure. This was possibly demonstrated by the two very low performing participants, who had large between test improvements and were subsequently omitted from analysis. Similarly, the TUG captured a broad range of values (Fig. 1b) and had excellent test-retest reliability, but there is no cutoff score that predicts high falls risk in adults with CP. Others have shown that the TUG is moderately associated with the 6-minute walk distance – explaining 67% of the variance in adults with CP [25]. This suggests that the TUG could be a useful measure of both balance and walking in adults with CP, but while the TUG MDC_{95} value can identify true changes in measured performance, the clinical implications of a given change need to be considered for each case.

Comfortable and fast walking speeds had a broad distribution of scores with “good” and “excellent” test-retest reliability respectively. Compared to CGS, the higher test-retest consistency of the FGS (Fig. 3) may be attributed to participants’ interpretations of the instructions. FGS instructions requested maximal effort, which may have been interpreted more consistently compared with CGS instructions which asked participants to recall a feeling (comfortable walking speed). Higher test-retest consistency led to a lower FGS MDC_{95} (0.14 m/s), compared to the CGS MDC_{95} (0.26 m/s).

MDC values intend to reflect only random variations in performance. Step length values at both speeds demonstrated small systematic increase in Test 2. All between - test differences were normally distributed and the cause of the small systematic shift was not identified. While this violation alludes to the need to use the step length MDC values with caution, they can still help assess changes in step length since the magnitude of the systematic differences was small (20–30% of the MDC values). Further, these MDC values remain clinically reasonable and similar to those of people with stroke [26].

Participants demonstrated a broad range of scores on most outcome measures, which suggests that these measures are able to capture

[2] $SEM = SD \times \sqrt{(1 - ICC)}$

[3] $MDC = SEM \times 1.96 \times \sqrt{2}$

Table 2
Test-retest reliability and minimal detectable change values of clinical balance measures.

	Test 1 mean (SD)		Test 2 mean (SD)		Paired t-test p value	$ICC_{2,1}$ (95%CI)	SEM	MDC_{95}	
BESTest - Total Score%	66.8	(17.7)	68.0	(17.5)	0.051	0.99 (0.96, 0.99)	1.76	4.9	%
I. Biomechanical Constraints%	51.3	(18.2)	52.7	(18.0)	0.163	0.97 (0.93, 0.99)	3.13	8.7	%
II. Stability Limits/Verticality%	81.0	(9.6)	81.0	(9.0)	1.000	0.88 (0.73, 0.95)	3.23	9.0	%
III. Transitions/Anticipatory%	60.0	(22.0)	60.6	(22.2)	0.681	0.97 (0.92, 0.99)	3.83	10.6	%
IV. Reactive%	56.9	(27.0)	59.7	(24.8)	0.204	0.93 (0.84, 0.97)	6.87	19.0	%
V. Sensory Orientation%	78.3	(19.6)	79.7	(20.0)	0.530	0.89 (0.75, 0.96)	6.56	18.2	%
VI. Stability in Gait%	69.5	(22.4)	71.0	(23.2)	0.186	0.98 (0.95, 0.99)	3.22	8.9	%
FRT (cm)									
Forward	20.3	(8.2)	21.3	(8.5)	0.238	0.90 (0.78, 0.96)	2.65	7.3	cm
Longer reach side	19.6	(6.2)	20.1	(6.3)	0.640	0.78 (0.54, 0.91)	2.93	8.1	cm
Shorter reach side	16.4	(5.6)	16.7	(5.7)	0.769	0.78 (0.53, 0.91)	2.63	7.3	cm
TUG (s)	12.2	(6.5)	11.5	(5.6)	0.065	0.97 (0.91, 0.99)	1.05	2.9	s
FSST (s)	14.8	(4.7)	14.0	(4.3)	0.088	0.91 (0.76, 0.96)	1.35	3.7	s
ABC %	68.3	(17.9)	68.7	(16.1)	0.880	0.86 (0.68, 0.94)	6.53	18.0	%
MFES	8.6	(1.1)	8.6	(1.1)	0.512	0.90 (0.78, 0.96)	0.36	1.0	

Abbreviations: FRT functional reach test; TUG Timed Up and Go; FSST Four Square Step Test; ABC Activities-specific Balance Confidence Scale; MFES, Modified Falls Efficacy Scale; ICC, intraclass correlation coefficient; CI, confidence interval; SEM, standard error of measurement; MDC, minimal detectable change.

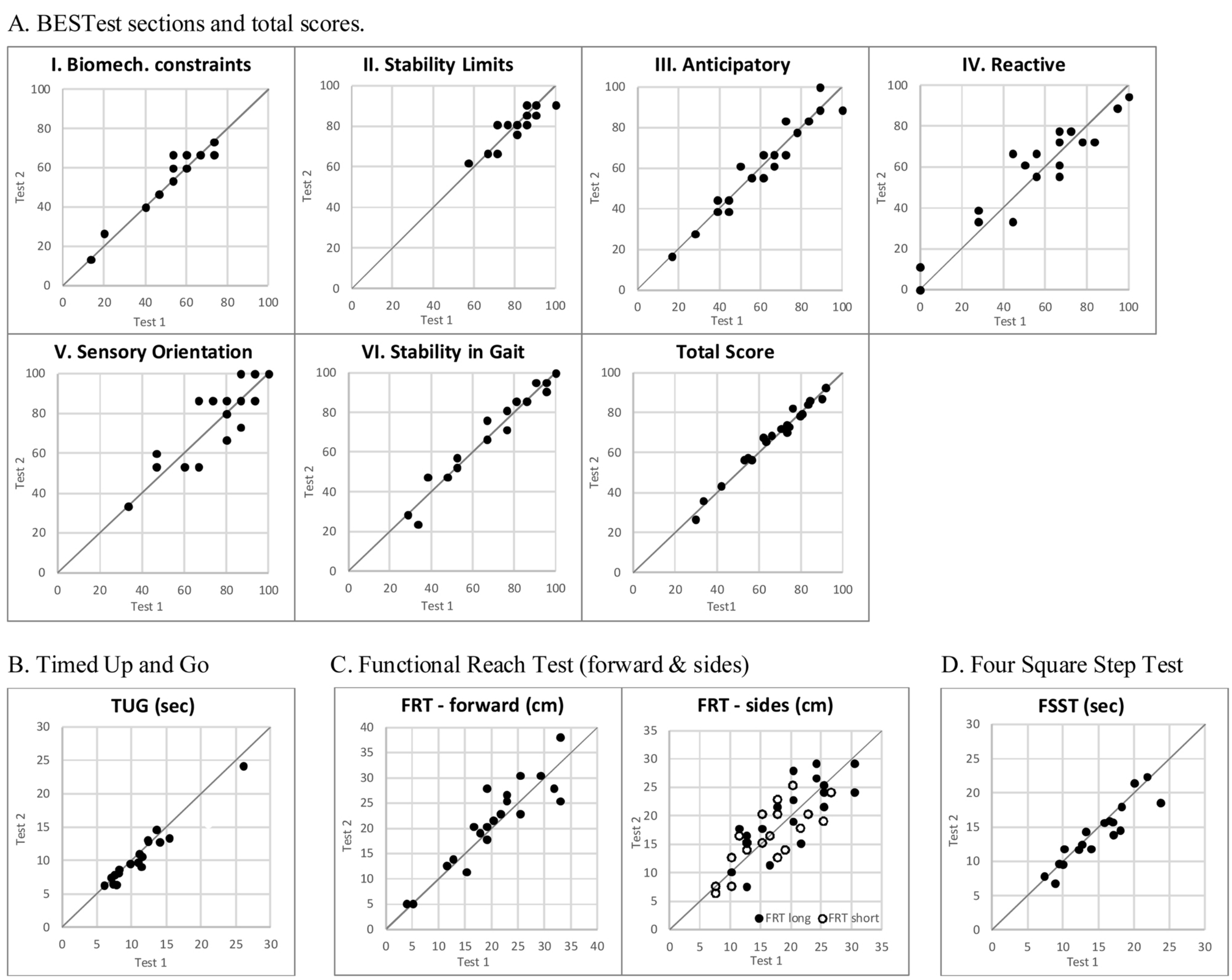

Fig. 1. Unity line score plots for balance outcome measures (Test 1 - Test 2). Dots on the unity line represent identical test-retest scores. Lower scores on Test 1 appear above the line and lower scores on Test 2 beneath the line.

population-specific functional range and have the potential to reflect changes in these functional ranges (Figs. 1–3). On BESTest Stability Limits and Sensory Orientation (Fig. 1a), and the MFES (Fig. 2) participants demonstrated score clustering close to the maximal score. This clustering reduces potential responsiveness to change and may result in a ceiling effect. To increase responsiveness to change in adults with CP with high "Stability Limits" scores, clinicians may consider the distance scores of the FRT items, evaluating changes by using FRT MDC_{95} values, and test those with high "Sensory Orientation" abilities with their feet together, as in the original instructions. Although the BESTest was found to have no floor or ceiling effects, and to be sensitive to balance changes in people with subacute stroke, 40% of which became

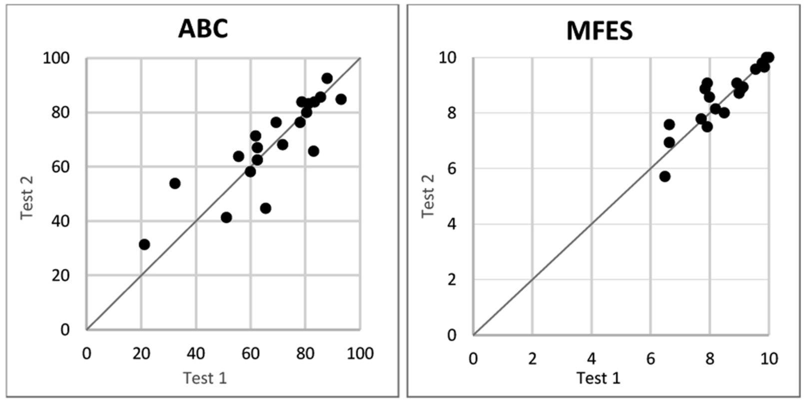

Fig. 2. Unity line score plots for Activities-specific Balance Confidence (ABC) and Modified Falls Efficacy Scale (MFES) questionnaires (Test 1 - Test 2). Dots on the unity line represent identical test-retest scores. Lower scores on Test 1 appear above the line, and lower scores on Test 2 beneath the line.

Table 3
Test-retest reliability and minimal detectable change values of gait variables.

		Test 1 mean (SD)		Test 2 mean (SD)		Paired t-test p value	$ICC_{2,1}$ (95% CI)	SEM	MDC_{95}	
Gait Velocity (m/s)	CGS	1.00	(0.27)	1.04	(0.28)	0.128	0.88 (0.73, 0.95)	0.10	0.26	m/s
	FGS	1.30	(0.36)	1.32	(0.36)	0.102	0.98 (0.94, 0.99)	0.05	0.14	m/s
Cadence (steps/min)	CGS	108.12	(16.42)	108.51	(15.40)	0.810	0.90 (0.78, 0.96)	5.03	14.00	steps/min
	FGS	123.35	(17.30)	122.98	(15.67)	0.775	0.94 (0.87, 0.98)	4.04	11.00	steps/min
Stride Length (m)	CGS	1.09	(0.21)	1.14	(0.22)	0.012*	0.92 (0.75, 0.97)	6.02	0.17	m
	FGS	1.24	(0.27)	1.27	(0.28)	0.010*	0.98 (0.92, 0.99)	3.88	0.11	m
Stride Width (m)	CGS	0.15	(0.07)	0.15	(0.06)	0.584	0.98 (0.94, 0.99)	0.95	0.03	m
	FGS	0.15	(0.07)	0.15	(0.06)	0.710	0.96 (0.91, 0.99)	1.31	0.04	m
Step Length - longer step (m)**	CGS	0.56	(0.10)	0.59	(0.10)	0.015*	0.90 (0.70, 0.96)	3.19	0.09	m
	FGS	0.64	(0.13)	0.66	(0.13)	0.006*	0.96 (0.85, 0.99)	2.60	0.07	m
Step Length - shorter step (m)**	CGS	0.53	(0.11)	0.55	(0.12)	0.019*	0.93 (0.79, 0.97)	3.04	0.08	m
	FGS	0.60	(0.15)	0.61	(0.15)	0.036*	0.99 (0.97, 1.00)	1.47	0.04	m
Double Support %	CGS	33.10	(4.35)	33.02	(8.14)	0.873	0.96 (0.90, 0.98)	1.55	4.30	%
	FGS	29.17	(6.53)	29.20	(6.86)	0.954	0.95 (0.88, 0.98)	1.50	4.20	%
Swing % - longer swing time**	CGS	35.02	(3.56)	34.94	(4.11)	0.795	0.94 (0.87, 0.98)	0.94	2.60	%
	FGS	37.02	(2.84)	36.95	(3.34)	0.820	0.92 (0.81, 0.97)	0.88	2.40	%
Swing % - shorter swing time**	CGS	31.86	(4.42)	32.01	(4.46)	0.585	0.96 (0.91, 0.99)	0.89	2.50	%
	FGS	33.68	(4.43)	33.71	(3.91)	0.913	0.95 (0.88, 0.98)	0.92	2.60	%

Abbreviations: CGS, comfortable gait speed; FGS, fast gait speed; ICC, intraclass correlation coefficient; CI, confidence interval; SEM, standard.
error of measurementMDC, minimal detectable change.
* Paired *t*-test p-value < 0.05, which suggest true difference between Test 1 and Test 2 means.
** Due to prevalent asymmetry in step lengths and swing time, steps on each side were defined, for each participant, as the "longer" or the "shorter".

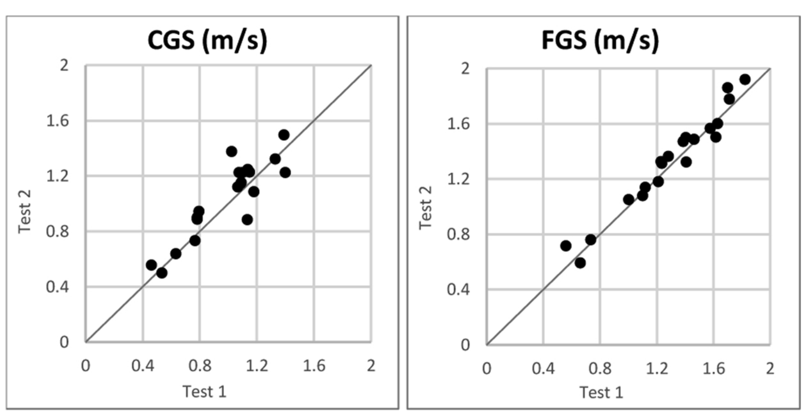

Fig. 3. Unity line score plots for comfortable gait speed (CGS) and fast gait speed (FGS) (Test 1 - Test 2). Dots on the unity line represent identical test-retest walking speeds. Lower walking speeds on Test 1 appear above the line, and lower walking speeds on Test 2 beneath the line.

independent ambulators [27], a structural validity study suggested that 25 out of the 36 items were the most appropriate to capture balance in ambulatory patients post stroke [28]. Revision of the BESTest specific to ambulant adults with CP may be appropriate.

The high scores clustering on the MFES was likely because the MFES does not include enough situations that are sufficiently challenging the balance of adults with CP. The Fall Efficacy Scale International (FES-I) [29] includes more challenging tasks and may better reflect the circumstances in which adults with CP experience falls, such as walking on uneven or slippery surfaces, walking around the neighborhood, and walking in crowds. For this study, the MFES was chosen over FES-I, because it uses a similar scoring scale as the ABC, while the FES-I is using a 4-point response scale. Future studies should evaluate test-retest reliability and the spread of adults with CP scores over the FES-I scale.

ABC score plot along the unity line, suggests that test-retest score consistency may depend on the functional level of the individual. The ABC questionnaire, which asks about confidence completing tasks without losing balance, had a “good” overall reliability (ICC = 0.86). While there was no significant difference between Test 2 and Test 1, there was a large confidence interval (CI) for the difference (Table 2) and the score plot (Fig. 2) demonstrated that lower scores tended to be further away from the unity line compared to the higher scores. The larger inconsistency in scores between tests, among lower functioning individuals, may reflect either fluctuations in confidence experienced by participants, or the inability to assign the same number consistently on an eleven-point scale. More consistent test-retest responses may be achieved by using a measure with fewer response options [29].

4.1. Limitations

There are several factors that limit the generalizability of these findings. The sample of convenience was not representative of the entire population of adults with CP GMFCS-E&R Levels I and II. The study only included participants that were able to follow instructions in English and no adults with CP who presented with significant communication or cognitive impairment were included, though no formal screening tools were used to assess communication and cognition.

5. Conclusions

A decline in balance and walking is experienced by many ambulant adults with CP. Reliable outcome measures are needed to support the research and services that address this decline. The current study established test-retest reliability and MDC values in ambulant adults with CP for outcome measures of balance and gait that were previously validated for other populations with balance impairments. Results show that these measures are useful for evaluating balance and gait in most ambulant adults with CP. Further studies are needed to improve the precision and responsiveness of some of these measures.

COI statement

The authors report no conflict of interest.

Declarations of interest

None.

Acknowledgement

This work was supported by Academy of Pediatric Physical Therapy, of the American Physical Therapy Association, Alexandria, VA, USA.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.gaitpost.2019.05.028.

References

[1] P. Rosenbaum, N. Paneth, A. Leviton, M. Goldstein, M. Bax, D. Damiano, B. Dan, B. Jacobsson, A report: the definition and classification of cerebral palsy April 2006, Dev. Med. Child Neurol. Suppl. 109 (2007) 8–14.

[2] P. Morgan, J. McGinley, Gait function and decline in adults with cerebral palsy: a systematic review, Disabil. Rehabil. 36 (2014) 1–9, https://doi.org/10.3109/09638288.2013.775359.

[3] D. Strauss, K. Ojdana, R. Shavelle, L. Rosenbloom, Decline in function and life expectancy of older persons with cerebral palsy, NeuroRehabilitation 19 (2004) 69–78.

[4] P. Haak, M. Lenski, M.J.C. Hidecker, M. Li, N. Paneth, Cerebral palsy and aging, Dev. Med. Child Neurol. 51 (Suppl 4) (2009) 16–23, https://doi.org/10.1111/j.1469-8749.2009.03428.x.

[5] A. Opheim, R. Jahnsen, E. Olsson, J.K. Stanghelle, Walking function, pain, and fatigue in adults with cerebral palsy: a 7-year follow-up study, Dev. Med. Child Neurol. 51 (2009) 381–388, https://doi.org/10.1111/j.1469-8749.2008.03250.x.

[6] I. Jeglinsky, J. Surakka, E.B. Carlberg, I. Autti-Rämö, Evidence on physiotherapeutic interventions for adults with cerebral palsy is sparse. A systematic review, Clin. Rehabil. 24 (2010) 771–788, https://doi.org/10.1177/0269215510367969.

[7] L.L. Tosi, N. Maher, D.W. Moore, M. Goldstein, M.L. Aisen, Adults with cerebral palsy: a workshop to define the challenges of treating and preventing secondary musculoskeletal and neuromuscular complications in this rapidly growing population, Dev. Med. Child Neurol. 51 (Suppl 4) (2009) 2–11, https://doi.org/10.1111/j.1469-8749.2009.03462.x.

[8] S.M. Haley, M.A. Fragala-Pinkham, Interpreting change scores of tests and measures used in physical therapy, Phys. Ther. 86 (2006) 735–743.

[9] F.B. Horak, D.M. Wrisley, J. Frank, The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits, Phys. Ther. 89 (2009) 484–498, https://doi.org/10.2522/ptj.20080071.

[10] P. Morgan, A. Murphy, A. Opheim, J. McGinley, Gait characteristics, balance performance and falls in ambulant adults with cerebral palsy: an observational study, Gait Posture 48 (2016) 243–248, https://doi.org/10.1016/j.gaitpost.2016.06.015.

[11] A. Opheim, R. Jahnsen, E. Olsson, J.K. Stanghelle, Balance in relation to walking deterioration in adults with spastic bilateral cerebral palsy, Phys. Ther. 92 (2012) 279–288, https://doi.org/10.2522/ptj.20100432.

[12] G. Leizerowitz, M. Katz-Leurer, Feasibility, stability and validity of the four square step test in typically developed children and children with brain damage, Brain Inj. 31 (2017) 1356–1361, https://doi.org/10.1080/02699052.2017.1332384.

[13] R.J. Palisano, P. Rosenbaum, D. Bartlett, M.H. Livingston, Content validity of the expanded and revised Gross Motor Function Classification System, Dev. Med. Child Neurol. 50 (2008) 744–750, https://doi.org/10.1111/j.1469-8749.2008.03089.x.

[14] BESTest: Training, (n.d.). http://www.bestest.us/training/ (accessed March 15, 2018).

[15] D.E. Thorpe, S.C. Dusing, C.G. Moore, Repeatability of temporospatial gait measures in children using the GAITRite electronic walkway, Arch. Phys. Med. Rehabil. 86 (2005) 2342–2346, https://doi.org/10.1016/j.apmr.2005.07.301.

[16] M.D. Lewek, E.P. Randall, Reliability of spatiotemporal asymmetry during overground walking for individuals following chronic stroke, J. Neurol. Phys. Ther. JNPT 35 (2011) 116–121, https://doi.org/10.1097/NPT.0b013e318227fe70.

[17] L.E. Powell, A.M. Myers, The Activities-specific Balance Confidence (ABC) Scale, J. Gerontol. A Biol. Sci. Med. Sci. 50A (1995) M28–34.

[18] K.D. Hill, J.A. Schwarz, A.J. Kalogeropoulos, S.J. Gibson, Fear of falling revisited, Arch. Phys. Med. Rehabil. 77 (1996) 1025–1029.

[19] W. Dite, V.A. Temple, A clinical test of stepping and change of direction to identify multiple falling older adults, Arch. Phys. Med. Rehabil. 83 (2002) 1566–1571.

[20] S.D. Walter, M. Eliasziw, A. Donner, Sample size and optimal designs for reliability studies, Stat. Med. 17 (1998) 101–110.

[21] J.E. Lexell, D.Y. Downham, How to assess the reliability of measurements in rehabilitation, Am. J. Phys. Med. Rehabil. Assoc. Acad. Physiatr. 84 (2005) 719–723.

[22] T.K. Koo, M.Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, J. Chiropr. Med. 15 (2016) 155–163, https://doi.org/10.1016/j.jcm.2016.02.012.

[23] Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2008http://www.R-project.org.

[24] W.G. Hopkins, Measures of reliability in sports medicine and science, Sports Med. 30 (2000) 1–15, https://doi.org/10.2165/00007256-200030010-00001.

[25] G. Maanum, R. Jahnsen, K.F. Frøslie, K.L. Larsen, A. Keller, Walking ability and predictors of performance on the 6-minute walk test in adults with spastic cerebral palsy, Dev. Med. Child Neurol. 52 (2010) e126–132, https://doi.org/10.1111/j.1469-8749.2010.03614.x.

[26] T.M. Kesar, S.A. Binder-Macleod, G.E. Hicks, D.S. Reisman, Minimal detectable change for gait variables collected during treadmill walking in individuals post-stroke, Gait Posture 33 (2011) 314–317, https://doi.org/10.1016/j.gaitpost.2010.11.024.

[27] B. Chinsongkram, N. Chaikeeree, V. Saengsirisuwan, N. Viriyatharakij, F.B. Horak, R. Boonsinsukh, Reliability and validity of the balance evaluation systems test (BESTest) in people with subacute stroke, Phys. Ther. (2014), https://doi.org/10.2522/ptj.20130558.

[28] K. Miyata, S. Hasegawa, H. Iwamoto, T. Otani, Y. Kaizu, T. Shinohara, S. Usuda, Structural validity of Balance Evaluation Systems Test assessed using factor and Rasch analyses in patients with stroke, J. Phys. Ther. Sci. 30 (2018) 1446–1454, https://doi.org/10.1589/jpts.30.1446.

[29] L. Yardley, N. Beyer, K. Hauer, G. Kempen, C. Piot-Ziegler, C. Todd, Development and initial validation of the Falls Efficacy Scale-International (FES-I), Age Ageing 34 (2005) 614–619.