https://doi.org/10.1016/j.exger.2018.11.001

RELIABILITY, VALIDITY AND MINIMAL DETECTABLE CHANGE OF 2-MINUTE WALK TEST, 6-MINUTE WALK TEST AND 10-METER WALK TEST IN FRAIL OLDER ADULTS WITH DEMENTIA

AUTHOR: CHAN, WAYNE L.S., PhD[1]; PIN, TAMIS W., PhD[2]

AFFILIATION: [1]PHYSIOTHERAPY DEPARTMENT, CHI LIN NUNNERY ELDERLY SERVICE, 5 CHI LIN DRIVE, DIAMOND HILL, HONG KONG; [2]DEPARTMENT OF REHABILITATION SCIENCES, THE HONG KONG POLYTECHNIC UNIVERSITY, HUNG HOM, HONG KONG

CORRESPONDING AUTHOR:

Name: CHAN, WAYNE L.S.

Address: Physiotherapy Department, Chi Lin Nunnery Elderly Service, 5 Chi Lin Drive, Diamond Hill, Kowloon, Hong Kong

Business phone number: +852 23541707

Email: waynels.chan@gmail.com

OTHER AUTHORS:

Name: PIN, TAMIS W.

Email: tamis.pin@polyu.edu.hk

DECLARATIONS OF INTEREST: None.

ABSTRACT

Background: Walk tests are commonly used to evaluate walking ability in frail older adults with dementia but their psychometric evidence in this population is lacking.

Objectives: 1) To examine test-retest and inter-rater reliability, construct and known-group validity, and minimal detectable change at 95% level of confidence (MDC_{95}) of walk tests in frail older adults with dementia, and 2) to examine the feasibility and consistency of a cueing system in facilitating participants in completing walk tests.

Design: Psychometric study with repeated measures.

Setting: Day care and residential care facilities.

Participants: Thirty-nine frail older adults with a mean age 87.1 and a diagnosis of dementia or Alzheimer's disease who were able to walk independently for at least 15 meters.

Methods: The participants underwent a 2-minute walk test (2MWT), 6-minute walk test (6MWT) and 10-meter walk test (10MeWT) on six separate occasions under 2 independent assessors using a cueing system. Functional status was measured using the Elderly Mobility Scale (EMS), Berg Balance Scale (BBS) and Modified Barthel Index (MBI).

Results: Excellent test-retest (ICC=.91-.98) and inter-rater reliability (ICC=.86-.96) were shown in the 2MWT, 6MWT and 10MeWT. The walk tests were strongly correlated with each other (ρ =.85-.94). The correlations between the walk tests and the functional measures were moderate in general (ρ =.34-.55). All the walk tests were able to distinguish between those who could walk outdoor and indoor only ($p \leq .036$). The MDC_{95} were 9.1m in the 2MWT, 28.1m in the 6MWT, and .16m/s in the 10MeWT. The cues provided by the assessors in the walk tests were generally consistent (ICC=.62-.89).

Conclusions: The 2MWT, 6MWT and 10MeWT are reliable and valid measures in evaluating walking ability in frail older adults with dementia. The MDC_{95} of the walk tests has been established. The cueing system is feasible and reliable to facilitate the administration of the walk tests in this population group.

Keywords: Dementia; Physical assessment; Psychometrics; Walking; Mobility

1. INTRODUCTION

Walking ability has been an important indicator of both physical and cognitive health among older adults who are at risk of developing dementia. Past studies have shown that reduced walking ability is associated with increased risk of developing cognitive impairment and functional decline in older populations (Blankevoort et al., 2013; Fitzpatrick et al., 2007; Jabourian et al., 2014; Taylor et al., 2014; Watson et al., 2010). Although assessing walking ability has been a part of routine assessment for older adults with dementia in many clinical settings, the psychometric properties of various walk tests have not yet been thoroughly examined for this population group (Fox et al., 2016). Exploring the psychometric properties, particularly reliability and validity, of walk tests would enhance the scientific rigor of their clinical use for older adults with dementia.

Many clinical-friendly walk tests, such as the 2-minute walk test (2MWT) (Butland et al., 1982), 6-minute walk test (6MWT) (Butland et al., 1982) and 10-meter walk test (10MeWT) (Collen et al., 1990), have been validated in older adults with normal cognition (Brooks et al., 2006; Harada et al., 1999; Hollman et al., 2008; Peters et al., 2013; Rikli and Jones, 1999; Steffen et al., 2002). However, only a few studies have explored the reliability of the 6MWT for older adults with dementia (Ries et al., 2009; Tappen et al., 1997). No study has investigated the reliability of the 2MWT and 10MeWT, and the validity of any walk test for people with dementia. Moreover, the two studies on the 6MWT recruited participants with unknown level of disability or functional independence. None of them specifically recruited frail older adults, who have reduced physiological capacity to recover from health stressors and increased risk of having disability and premature death (Morley et al., 2012). Looking into all types of walk tests in general, the findings of psychometric studies on older adults with dementia have been inconsistent. For instance, while the test-retest reliability of walking speed measures has been shown acceptable in some studies (Ries et al., 2009; Thomas and Hageman, 2002; van Iersel et al., 2008), some studies demonstrated poor reliability (Fox et al., 2014; Tappen et al., 1997). The psychometric properties of walk tests remain unknown for this population group.

Assessing walking ability accurately in older adults with dementia has been a challenge for many clinicians. Cognitive impairment has been identified as the key determinant of inconsistent measurement of any physical performance test, and the short attention span and inadequate compliance to instructions found in people with dementia limit the use of walk tests in clinical settings (Rockwood et al., 2000). Previous studies have documented that frequent verbal and physical assistance and radical changes to the test protocols were required to ensure the successful completion of performance assessments, including walk tests, for this population group (Maring et al., 2013; Nordin et al., 2006; Ries et al., 2009; Tappen et al., 1997). Recently, a study has validated a cueing system to guide clinicians to conduct performance tests on people with dementia (Ries et al., 2009). Nevertheless, the feasibility of using this cueing system to monitor the assistance offered to people with dementia and to facilitate clinicians how to provide appropriate verbal and physical prompts during walk tests remain unknown.

To fill in the present research gap in walk tests for older adults with dementia, this study investigated the test-retest and inter-rater reliability, construct and known-group validity, and minimal detectable change (MDC) of the 2MWT, 6MWT and 10MeWT for this population group. In addition, a progressive cueing system was used to facilitate the participants to complete the walk tests. The consistency of the assessors in providing repeated cues in the walk tests was also examined.

2. METHODS

2.1. Study design

This psychometric study is a cross-sectional, non-experimental study with repeated measures. To examine the test-retest reliability of the walk tests, the participants completed the 2MWT, 6MWT and 10MeWT with one assessor on two test occasions. For the inter-rater reliability, the participants repeated the walk tests with another assessor on a separate test occasion. The construct validity was examined based on the correlations between the walking performances and functional measures commonly applied in older populations, including the Berg Balance Scale (BBS), Elderly Mobility Scale (EMS) and Modified Barthel Index (MBI) which reflected the balance control, general mobility and daily functioning of the participants respectively. The known-group validity was assessed by comparing the walking performances among participants using different walking aids (no aid vs stick vs quadripod/frame) and those having different ambulatory statuses (indoor vs outdoor walking). The MDC of the walk tests were determined based on the findings of the test-retest reliability and the walking performances using standardized formula. Lastly, the test-retest and inter-rater reliability of the progressive cueing system was explored to assess the consistency of the assessors in providing repeated cues in the walk tests.

2.2. Participants

Participants were recruited from January to May 2016 from a day care center and a residential care facility providing permanent medical care services to older adults with moderate to severe disabilities. Individuals who were: 1) 65 years or above; 2) able to walk 15 meters independently with or without walking aids; 3) diagnosed with dementia or Alzheimer's disease; 4) those without a proper diagnosis of dementia or Alzheimer's disease but scored below the cut-off point of 19 in the Chinese Mini-Mental State Examination (CMMSE) (Chiu et al., 1994); and 5) scored 3 or above in the FRAIL scale (Morley et al., 2012; Woo et al., 2015) were eligible to join the study. Older adults with acute exacerbation of cardiac, pulmonary or musculoskeletal conditions that affected their walking ability, those with severe hearing or visual impairment that hindered effective communication, and those with recent hospitalization in the past 30 days were excluded. Potential participants were identified by the health care professionals working in the facilities. Then they were referred to an in-house physical therapist (the first author) to screen for their walking capacities and medical conditions to make sure that they could complete the walk tests safely. Figure 1 shows the recruitment process of this study.

2.3. Sample size calculation

Based on the limited psychometric studies on community-dwelling, healthy individuals with dementia, the findings on the test-retest reliability of walk tests were generally favorable (Intra-class Correlation Coefficient (ICC) ≥ .92) (Ries et al., 2009; Tappen et al., 1997; Thomas and Hageman, 2002). Assuming an ICC ≥ .90 indicates strong reliability, a sample size of 30 was required to achieve 90% power at a confidence level of 0.05 (Shoukri et al., 2004). Anticipating a 20% of unexpected drop-outs, this study aimed at recruiting 38 participants.

2.4. Procedures

The study was complied with the Declaration of Helsinki and approved by the ethics committees of the Hong Kong Polytechnic University and the participating facilities in order to gain access to the medical records of the participants and conduct the study on the participants. Eligible participants, and their family members or guardians were provided with written information about the proposed study, including the objectives, tests and duration of the study face-to-face or through mail. Family members or guardians of the participants signed the informed written consent for the participants by proxy.

Demographic data, including age, gender, height, weight, body mass index and past medical history, which were reviewed yearly by the health care professionals in the facilities, were assembled from the medical records of the participants. The scores of functional measures, including the Berg Balance Scale (BBS), Elderly Mobility Scale (EMS) and Modified Barthel Index (MBI), and ambulatory status were assessed by the in-house physical therapist (the first author) during the study period. The ambulatory status of the participants were categorized into either indoor or outdoor walkers according to the Modified Functional Ambulation Classification (Chau et al., 2013). Indoor walkers can transfer, turn and walk independently on levelled ground but requires supervision or physical assistance to ambulate on stairs, incline or uneven surface.

Outdoor walkers can ambulate independently in any kind of surface or stairs (Chau et al., 2013). Table 1 shows the descriptive characteristics of the participants.

The walk tests were performed from April to July 2016. Two experienced physical therapists (the first author and a physical therapist of the day care center) conducted the walk tests independently. Over a two-week period, each participant underwent the 2MWT, 6MWT and 10MeWT on six separate occasions (figure 1). Assessor A (the first author) conducted the walk tests on four occasions so as to examine the test-retest reliability of the walk tests, while Assessor B (the physical therapist of the day care center) performed the tests on two occasions aiming to examine the inter-rater reliability of the walk tests. The measurement occasions were at least one day apart, and the participants were given adequate rest before proceeding to another occasion on the next day. Any acute change to their medical conditions during the 2-week period were noticed by the assessors and the participant concerned would be excluded from the study. To prevent bias and minimize potential learning effect, the sequence of the test occasions was randomized by drawing lots and both assessors were blinded to previous test performances.

2.5. Measures

A 15-meter levelled corridor with colored markings at every 1-meter interval was assigned for the walk tests. Traffic cones were placed to indicate the turning spots at both ends of the corridor. The participants were asked to wear comfortable clothing and use their usual walking aids. No vigorous exercise was allowed two hours before the tests started. Heart rate, blood pressure and pulse oxygen saturation were recorded using a blood pressure monitor and finger pulse oximetry before and after each test. The participants were instructed to report to the assessor and stop the test if they experienced any discomfort, including dizziness, chest pain, nausea and undue fatigue. The rate of perceived exertion was not monitored and recorded in our study as required by the published guidelines (Crapo et al., 2002) as most of the participants, who had moderate to severe grade of dementia or Alzheimer's disease, had difficulty in comprehending and grading their perceived exertion. All the vital signs had to return to the baseline before the next trial commenced (Crapo et al., 2002; Pin, 2014).

Strategies were implemented to maximize the attention span of the participants during the walk tests as follows (Hoppes et al., 2003; Kovach and Henschel, 1996; Miller, 2008). The corridor was located at a quiet, spacious hall inside the participating facilities. Other staff members and people in the facilities were prohibited from entering the venue during the tests. The tests were performed by the therapists working in the facilities whom the participants were familiar with. The therapists were instructed to build an effective interaction with the participants, such as using friendly, pleasant voice and facial expressions, providing clear commands, and keeping constant eye contact (Miller, 2008; Small et al., 2003). The tests were conducted at about the same time of the day for each participant.

2.5.1. 2-minute Walk Test (2MWT)

The 2MWT was conducted based on the published guideline (Pin, 2014). The participants were instructed to "walk at your comfortable, usual pace". Two practice trials and a final trial for record were performed. The two practice trials were used to minimize a possible learning effect

(Pin, 2014). At least ten minutes of rest was provided between trials. The distance covered in the two minutes was recorded as the 2MWT.

2.5.2. 6-minute Walk Test (6MWT)

The 6MWT was performed following the published guideline (Crapo et al., 2002). The participants were asked to walk "as far as possible" in the six minutes. No running was allowed. The assessors provided standardized encouragement, "you're doing well, keep it up", every minute during the test. One practice trial and one final trial for record were performed. The practice trial was used to make sure that the participants were medically safe and physically capable to complete the test on any particular occasion. A minimum 20 minutes of rest was given between trials. The distance covered in the 6 minutes was recorded as the 6MWT.

2.5.3. 10-meter Walk Test (10MeWT)

The 10MeWT was measured concurrently in the 2MWT (10MeWT-2M) and 6MWT (10MeWT-6M) to obtain the walking speeds of the participants. One reason for the simultaneous measurements was that the test protocols and the environmental set-up of the 2MWT, 6MWT and 10MeWT were very similar. Combining these tests could reduce the number of repeated walking and the resulting fatigue for the participants, and thus maximized their compliance to the tests. The walking speeds achieved in the 2MWT (10MeWT-2M) and 6MWT (10MeWT-6M) were treated as distinctive outcomes and analyzed separately because we believed that the instructions given to the participants in the 2MWT and 6MWT were different, possibly resulting in differences in the timed walk tests.

The time started when the participants walked for the middle 10 meters of the 15-meter corridor in the first leg of the recorded trial. The first three meters and the last two meters were reserved for acceleration and deceleration respectively (Flansbjer et al., 2005; Hollman et al., 2008; Peters et al., 2013). The walking speeds were calculated by dividing 10 meters by the time used (i.e. meters/second).

2.5.4. Berg Balance Scale (BBS)

The BBS assesses the balance control of older adults based on the performance of various functional tasks (Berg et al., 1992, 1989). Fourteen functional tasks, such as sitting to standing, standing unsupported, chair transfers, standing with eye closed, tandem standing and single leg standing. A scale from 0 to 4 is used for each item. A score of 0 indicates the inability to perform the task, while a score of 4 indicates the completion of the task successfully according to the predetermined criterion. The maximum score is 56. A higher score indicates better balance control. Moderate to strong test-retest and inter-rater reliability was reported among people with dementia (ICC=.72-.99) (Muir-Hunter et al., 2015; Telenius et al., 2015).

2.5.5. Elderly Mobility Scale (EMS)

The EMS assesses the general mobility, balance and position changes of frail older adults (Smith, 1994). Seven mobility and functional tasks, including lying to sitting, sitting to lying, sitting to standing, standing, gait, 6-meter timed walk and functional reach, were used. Two items, lying to sitting and sitting to lying, score from 0 to 2. Four tasks, including sitting to standing, standing, gait, 6-meter timed walk, have a score ranging from 0 to 3. Functional reach scores from 0 to 4. The total score ranges from 0 to 20. A higher score indicates better mobility status. Strong inter-rater reliability (Spearman correlation coefficient, ρ =.88) and strong correlation with the Barthel Index (ρ =.79) were reported among older adults (Prosser and Canby, 1997).

2.5.6. Modified Barthel Index (MBI)

The MBI measures the level of functional independence in activities of daily living (Leung et al., 2007; Shah et al., 1989). Ten tasks of activities of daily living, including feeding, transfer, personal hygiene, getting on/off toilet, bathing, walking on levelled ground, climbing stairs, dressing, bowel and bladder control, were evaluated. The performances of the participants were reported by the family members or caregivers of those who attended the day care center, and by the personal care workers and nurses of those living in the residential care facility. Direct observation was also conducted by the first author if necessary. Each item scores from 0 to 10, which adds up to a total score ranging from 0 to 100. A higher score indicates higher level of functional independence. The test-retest reliability has been shown moderate to strong (Kappa coefficients=.63-1.00) among local older adults (Leung et al., 2007).

2.6. The cueing system

The system aims to facilitate clinicians how to monitor and quantify the verbal and physical assistance and provide consistent assistance during functional assessment for people with dementia (Nordin et al., 2006; Ries et al., 2009; van Iersel et al., 2007). It was originally developed to assist the clinicians to make decision how to provide assistance in activities of daily living for people with cognitive impairment (Beck et al., 1993). The level of cueing is escalated in the following sequence: 0) no cue; 1) verbal prompt; 2) modelling/gesturing; 3) one-off physical prompt; 4) intermittent physical prompt; 5) intermittent physical guidance; and 6) complete physical guidance. In the present study, cues were given when the participant started to deviate from the walk path, to run or slow down, or to stop walking during the walk tests. The participant would be given a few seconds to respond to the cue before the next level of cue was given. For both levels 5 and 6 of cueing, the assessor would walk in front of the participant and held one of his/her hands intermittently or continuously respectively without pulling to guide the direction of the walking. The physical touch or the number of repeated cueing given was kept as minimal as possible to allow the participant to initiate and sustain their walking. The practice trials were used to determine which level of cueing should be provided to the participants. If walking ahead of the participant was required, the assessor would adopt the walking pace of the participant as observed in the practice trials to minimize the risk of driving the pace of walking. Figure 2 shows the flow diagram how the assessors determined the appropriate level of cueing for the participants.

2.7. Statistical analyses

Reliability indicates the degree to which scores of an outcome measure are free from measurement errors within a particular population (Mokkink et al., 2010). Reliability is further divided into relative reliability, which indicates the consistency of the ranking of individuals' scores within a group, and absolute reliability, which indicates the variability of individuals' scores in repeated measurements (Carter, R. E., Lubinsky, J., Domholdt, 2011). The test-retest and inter-rater reliability of the walk tests, the sub-categories of the relative reliability, were analyzed using the Intra-class Correlation model 2 ($ICC_{2,1}$) and model 3 ($ICC_{3,2}$) respectively (Portney and Watkins, 2000). Given the small sample size in the present study, Spearman correlation coefficient (ρ) was used to analyze the correlation between the walk tests and the other functional measures. A coefficient between .30 and .60 indicates moderate correlation, and $\rho \geq .60$ excellent correlation (Andresen, 2000; Fitzpatrick et al., 1998). For the construct validity, moderate correlation is commonly considered acceptable (Brooks et al., 2006). Known-group validity was examined by comparing the performance of the participants using different walking aids and of different ambulatory statuses using independent *t* test (two-tailed) or one-way analysis of variance (ANOVA).

Absolute reliability provides additional information about whether a change of performance is beyond expected measurement error and takes individual variabilities into account. Absolute reliability is evaluated using the minimal detectable change (MDC), which refers to the amount of change required to demonstrate a "true" change in the individual's performance (Beckerman et al., 2001; Weir, 2005). The MDC is expressed as an absolute value at the 95% confidence level (MDC_{95}), which depends on the units of measure, or a relative value in percentage ($MDC_{95}\%$), which determines a change from the baseline over time. The MDC_{95} indicates the smallest difference required to exceed the measurement error and performance variability with 95% confidence.

The calculation of the MDC was based on the results within the same assessor on the participants. Firstly, the standard error of measurement (SEM) was calculated using the following formula (Carter, R. E., Lubinsky, J., Domholdt, 2011):

$$SEM = sd \times \sqrt{(1 - r)}$$

where *sd* is the standard deviation of the measure, and *r* is the reliability coefficient, i.e. the ICC of the test-retest and inter-rater reliability.

The absolute (MDC_{95}) and the relative MDC at the 95% confidence level ($MDC_{95}\%$) were calculated based on the SEM. The MDC_{95} was calculated according to the following equation (Beckerman et al., 2001; Weir, 2005):

$$MDC_{95} = SEM \times 1.96 \times \sqrt{2}$$

where 1.96 represents the z-score at the 95% confidence interval from a normal distribution. The square root of 2 takes into account the errors made by the repeated measurements.

The $MDC_{95}\%$ was based on the following formula (Beckerman et al., 2001; Flansbjer et al., 2005):

$MDC_{95}\% = (MDC_{95}/mean) \times 100$

where "mean" was the average results of the walk tests.

To examine the consistency of the assessors in providing repeated cues, the ICC model 2 ($ICC_{2,1}$) and model 3 ($ICC_{3,2}$) were used to analyze the consistency of the repeated cues provided by the same assessor (i.e. test-retest reliability) and across two assessors (i.e. inter-rater reliability) respectively (Portney and Watkins, 2000).

The SPSS software (version 22.0) was used to perform all statistical analyses. A significance level of 0.05 was used for all analyses.

3. RESULTS

Thirty-seven participants with a diagnosis of dementia or Alzheimer's disease and two with no formal diagnosis of dementia but scored below the cut-off point in the CMMSE were recruited and completed all the tests without any adverse events. Table 1 shows the characteristics of the participants. Although all the metric data were normally distributed, with the present sample size, the statistical analysis methods remained as discussed above. The reliability coefficients of all the walk tests are shown in table 2. All the walk tests achieved an excellent test-retest

(ICC=.91-.98) and inter-rater reliability (ICC=.86-.96), except the 10MeWT-6M, which attained only moderate inter-rater reliability (ICC=.60-.65).

Table 3 shows the correlations among the walk tests and other functional measures. All the walk tests were strongly correlated with each other (ρ =.84-.93). Moderate correlations were found between the 2MWT, 6MWT and 10MeWT-2M and the EMS (ρ =.39-.43), BBS (ρ =.47-.49) and MBI (ρ =.45-.54). For the 10MeWT-6M, the correlations with the BBS and MBI were moderate (ρ =.35 and .46 respectively), but the one with the EMS (ρ =.27) was weak and insignificant.

Participants who ambulated without a walking aid covered significantly longer distance than those using any kinds of walking aids in both the 2MWT and 6MWT (all $p \leq .05$) (table 4). However, such differences were not found in the 10MeWT-2M and 10MeWT-6M (all $p \geq .05$). Participants who were able to walk outdoor outperformed those who walked indoor only in all the walk tests (all $p \leq .05$).

The SEM, MDC_{95} and MDC_{95}% of the walk tests are shown in table 5. The MDC_{95} of the 2MWT, 6MWT, 10MeWT-2M and 10MeWT-6M were 9.1m, 28.1m, .17m/s, and .16m/s respectively. The MDC_{95}% in the 2MWT, 6MWT, 10MeWT-2M and 10MeWT-6M were 14%, 14%, 29%, and 24% respectively.

Table 6 shows the medians and reliability coefficients of the cues provided by the assessors in the walk tests. Majority of the participants received either no cue (level 0) (2MWT and 10MeWT-2M: 25.6-41.0%; 6MWT and 10MeWT-6M: 20.5-43.6%) or verbal prompt (level 1) (2MWT and 10MeWT-2M: 35.9-48.7%; 6MWT and 10MeWT-6M: 28.2-41.0%). Very few participants received intermittent (level 5) (2MWT and 10MeWT-2M: 0.0-2.6%; 6MWT and 10MeWT-6M: 2.6-7.7%) and complete physical guidance (level 6) (2MWT and 10MeWT-2M: 2.6-5.1%; 6MWT and 10MeWT-6M: 0.0-5.1%). The test-retest reliability (ICC=.88-.89) and inter-rater reliability between Assessor A Occasion 1 and Assessor B were excellent (ICC=.83). The inter-rater reliability between Assessor A Occasion 2 and Assessor B were modest (ICC=.62-.69).

4. DISCUSSION

This study aimed to examine the test-retest and inter-rater reliability, construct and known-group validity, MDC_{95} and MDC_{95}% on the 2MWT, 6MWT, 10MeWT in older adults with dementia. It is the present research gap in using these walk tests for this challenging population, who have difficulties in following instructions and sustaining adequate attention during physical training and performance tests (Rockwood et al., 2000). The walking performances of our participants were generally worse than healthy older adults who participated in other psychometric studies (Bohannon et al., 2015; Peters et al., 2013; Steffen et al., 2002), indicating that there is a need to determine the psychometric properties of the walk tests particularly for this population group. To support people with dementia during the tests, we also investigated the use of a standardized and individualized cueing system for this population group during the walk tests. We found that the system was feasible and generally reliable among our participants.

4.1. 2MWT

The distance walked by our participants was shorter than that by healthy older adults living in the community (2MWT: 63.1-63.6m vs 134.3-184.2m) (Bohannon et al., 2015). Our study is the first study to explore the reliability and validity of the 2MWT in the older adults with dementia. A study of young elderly with intellectual developmental disability reported a high correlation between the 2MWT and MBI (*r*=.75) (Maring et al., 2013). The SEM and MDC_{95} found in the current study are smaller than those of the older adults with normal cognition (SEM: 3.3-6.2m vs 5.2-6.3m; MDC_{95}: 9.1-17.1m vs 14.5-17.5m) (Connelly et al., 2009). The previous study has a smaller sample size (n=16), a high dropout rate (n=9) and a wider dispersion of the 2MWT (*SD*= 23.3-25.6m) (Connelly et al., 2009), which are believed to contribute to the discrepancies between our and their findings.

4.2. 6MWT

The 6MWT achieved by our participants was less than community-dwelling healthy adults who aged 60-89 (6MWT: 194.0-203.6m vs 392-572m) (Steffen et al., 2002). Our findings are consistent with previous findings that the 6MWT had high test-retest and inter-rater reliability in older adults with dementia (Ries et al., 2009; Tappen et al., 1997). The current study further demonstrates that the 6MWT was reliable and significantly correlated with other walk tests and functional measures in frail elderly with dementia with high level of functional dependence. Apart from being a tool to assess walking performance, the 6MWT is also regarded as a clinical test to assess submaximal exercise capacity in people with limited cardiopulmonary reserve

(Crapo et al., 2002). The moderate correlations between the 6MWT and various functional measures indicates that the 6MWT is also valid in measuring exercise capacity of this population group. The SEM and MDC_{95} reported in this study were smaller than the previous study conducted in people with dementia (SEM:10.1-17.6m vs 19.6-21.9m; MDC_{95}:28.1-48.7m vs 54.2-60.6m) (Ries et al., 2009). Despite having a similar level of cognitive function (mean MMSE=13.1), the younger (mean age=80.7), more independent participants in that study (community-dwelling participants= 76.5%) would have contributed to the larger SEM and MDC (Ries et al., 2009).

4.3. 10MeWT

Our participants generally walked slower than their healthy counterparts participated in a previous study (10MeWT-2M: .62-.63m/s; 10MeWT-6M: .64-.65m/s vs .96m/s) (Peters et al., 2013). The reliability and validity of the 10MeWT in people with dementia have never been examined. The excellent test-retest reliability found in our study is consistent with the findings of previous studies conducted on older adults with normal cognition (Hollman et al., 2008; Peters et al., 2013), which suggests that testing the older adults with dementia on one occasion would be sufficient in clinical setting. The inter-rater reliability of the 10MeWT-2M was excellent, but the inter-rater reliability of the 10MeWT-6M was only modest (table 2). We speculate that the poor inter-rater reliability of the 10MeWT-6M was affected by the two outliers, who had extreme performances across the two occasions (Supplementary Figure 1). After eliminating these two outliers, the inter-rater reliability of the 10MeWT-6M became excellent (ICC=.87-.90, 95% CI=.77-.96). Hence, it is reasonable to conclude that the test-retest and inter-rater reliability of the 10MeWT for older adults with dementia was good to excellent.

The SEM and MDC_{95} of the 10MeWT in this study were generally larger than those of the older adults with normal cognition (SEM:.06m/s vs.004-.06m/s; $MDC_{95:}$.16-.17m/s vs.01-.17m/s) (Hollman et al., 2008; Perera et al., 2006; Peters et al., 2013). The better cognitive function of the participants in those studies may have contributed to the smaller SEM and MDC because of the more consistent performance in the participants, i.e. smaller SD of the findings.

We have identified several characteristics about these two outliers, who both had extreme walking speeds on the 2 occasions. Both outliers were highly mobile (EMS=18-20 vs group mean =17.4; BBS=46-54 vs group mean = 44.1) and could ambulate independently without any walking aid. However, they were more cognitively impaired than other participants (MMSE=6-12 vs group mean =13.2). Moreover, severe behavioral problems, such as loss of attention, sudden change of the walking speed and deviations from the walking path, were observed on both participants during the walk tests. This observation was coherent with a previous study which found that people with severe dementia had huge difficulty in following instructions, causing the poor reliability in some physical performance measures such as walking speed (Tappen et al., 1997). Whenever an individual with dementia loses his/her attention or has unexpected behavior during a performance test, such as walk tests, it takes time for them to respond to verbal or physical cues and resume the walking. The extra time required to respond to the cues will greatly affect the result of the performance test. In addition, these unexpected behaviors could hardly be predicted. All these factors would become an issue in performance tests, particular to those that require the participants to complete in a very short period of time, such as the 10MeWT in our study.

Based on our results, the 10MeWT was a reliable and valid measure and able to capture changes in the walking performance of older adults with dementia. However, clinicians have to pay attention to those, who are highly mobile and have severe cognitive impairment with significant behavioral problems, may show large variations in their walking performance. Strategies, such as using simple and clear instructions and asking them to use comfortable pace as in the 2MWT, may improve the repeatability of the 10MeWT. Standardized testing protocol of the 10MeWT for people with or without dementia is urgently required so that clinicians can confidently administer this walk test in clinical settings.

4.4. The progressive cueing system

Psychometric studies of performance measures on dementia populations are scarce because cognitive function often determines whether an individual is able to complete any physical performance test (Rockwood et al., 2000). Our study has demonstrated that the walk tests have good reliability and validity in older adults with dementia. We believe that the progressive cueing system has played an important role in demonstrating that older adults with dementia were capable to complete the walk tests when systematic cues were provided.

In order to ensure people with dementia could complete the walk test, previous studies commonly made lots of modifications to the testing protocols of performance measures, such as introducing a “pacer” to provide ongoing encouragement and cueing (Maring et al., 2013) and skipping the turn and transfer in the timed up-and-go test (Tappen et al., 1997). These modifications were sometimes so extreme that they had changed the construct of the actual tests. The present study, which adapted the cueing system (Ries et al., 2009), demonstrated that there was no need to make changes to the testing protocol if the cueing system was used. A low-stress environment that minimizes distraction and some simple strategies to prolong the attention of older adults, such as using the quiet and spacious venue and familiar personnel from the facilities as the assessors, (Nordin et al., 2006; Ries et al., 2009; Thomas and Hageman, 2002; van Iersel et al., 2007) were also proved to be helpful in the present study.

Some may argue that the reliability analyses examined the consistency of the cues provided by the assessors rather than the performance of the participants. Any physical performance test for individuals with dementia would become a measure of the performance of the assessor if the cues are not systematically provided, particularly when physical assistance is involved (Tappen et al., 1997). We analyzed the test-retest and inter-rater reliability of the cueing and found a different pattern of consistency compared to that of the walk tests. For example, the inter-rater reliability of the cueing in the 2MWT and 10MeWT-2M were generally smaller than those of the performance of the participants (ICC=.62-.83 vs .86-.96). The findings show that the consistency of the cueing and the performance of the participants were independent of each other. Therefore, the reliability analyses evaluated the performance of the participants, not the assessors. In fact, the cueing system facilitated the walk tests to reveal the true walking ability of the participants. Strategies were implemented to ensure consistent provision of cueing, including the use of practice trials and the explicit guidelines of the walk tests, strategies to minimize the use of any unnecessary cueing. We believe that the practice trials gave both the participants and the assessors time to be more familiar with the tests and the cueing system.

4.5. Study Limitations

Our participants were relatively old, frail, mostly female, and recruited from day care and residential care facilities. Our findings may not be generalized to younger, non-frail, community-dwelling older men with dementia. Although the number of participants recruited in our study was larger than previous psychometric studies of walk tests for older adults (Connelly et al., 2009; Maring et al., 2013; Tappen et al., 1997; Thomas and Hageman, 2002), the sample size remained small. Future studies with larger sample sizes are suggested so to allow sub-group comparisons such as different severity of dementia. Two participants without a diagnosis of Alzheimer's disease or dementia but with significant cognitive impairment were recruited in our study. More stringent inclusion criteria aiming at those with a formal diagnosis may be necessary. The length of the corridor used in our study was only 15 meters, although 30 meters (or not less than 20 meters) is suggested to avoid excessive turns at the end of the corridor (Crapo et al., 2002; Pin, 2014). The distance walked by our participants may be less than expected due to the shorter corridor and the increased number of turns. However, our participants were assessed under the same settings across multiple test occasions, the effects of variations in environmental factors on walking performance was thus reduced. Clinicians should be aware of the effects of different testing settings on their findings. Our assessors were not blinded to the study, which may create possible bias to the measurements. Furthermore, the 10MeWT-2M and 10MeWT-6M were measured within the 2MWT and 6MWT respectively in the present study to maximize the compliance of the participants to the measurements. The results might be different if the tests were conducted separately. The potential influence of the two outliers on the inter-rater reliability of the 10MeWT-6M was a speculation based on our findings and the observation of our assessors. Future studies aiming at people with moderate to severe dementia and behavioral symptoms are required to confirm our speculation.

5. CONCLUSION

The 2MWT, 6MWT and 10MeWT had excellent test-retest and inter-rater reliability in measuring the walking performance in frail older adults with dementia who were receiving permanent care in day care and residential care facilities. These walk tests were valid and significantly correlated with functional measures commonly conducted in older populations. The MDC_{95} and $MDC_{95}\%$ of the walk tests were recommended for this population group. The progressive cueing system could facilitate the frail older adults with dementia to complete the walk tests with good reliability. Future studies on younger, non-frail and community-dwelling individuals should be conducted to expand the application of these walk tests among people with dementia.

FUNDING:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ACKNOWLEDGEMENT:

We would like to express our gratitude to Dr. Raymond Chung for his support on the data analyses. We would like to thank all the older adults who participated in the study. Special thanks are also given to the staff in the Chi Lin Nunnery Elderly Service for their support in the study.

REFERENCES

Andresen, E.M., 2000. Criteria for assessing the tools of disability outcomes research. Arch. Phys. Med. Rehabil. 81, 15–20. https://doi.org/10.1053/apmr.2000.20619

Beck, C., Heacock, P., Rapp, C.G., Mercer, S.O., 1993. Assisting cognitively impaired elders with activities of daily living. Am. J. Alzheimers. Dis. Other Demen. 8, 11–20. https://doi.org/10.1177/153331759300800602

Beckerman, H., Roebroeck, M.E., Lankhorst, G.J., Becher, J.G., Bezemer, P.D., Verbeek, A.L.M., 2001. Smallest real difference, a link between reproducibility and responsiveness. Qual. Life Res. 10, 571–578. https://doi.org/10.1023/A:1013138911638

Berg, K.O., Wood-Dauphinee, S.L., Williams, J.I., Gayton, D., 1989. Measuring balance in elderly: preliminary development of an instrument. Physiother. Canada 41, 304–311.

Berg, K.O., Wood-Dauphinee, S.L., Williams, J.I., Maki, B., 1992. Measuring balance in the elderly: validation of an instrument. Can. J. Public Health 83 Suppl 2, S7-11.

Blankevoort, C.G., Scherder, E.J., Wieling, M.B., Hortobagyi, T., Brouwer, W.H., Geuze, R.H., van Heuvelen, M.J., 2013. Physical predictors of cognitive performance in healthy older adults: a cross-sectional analysis. PLoS One 8, e70799. https://doi.org/10.1371/journal.pone.0070799

Bohannon, R.W., Wang, Y.C., Gershon, R.C., 2015. Two-minute walk test performance by adults 18 to 85 years: Normative values, reliability, and responsiveness. Arch. Phys. Med. Rehabil. 96, 472–477. https://doi.org/10.1016/j.apmr.2014.10.006

Brooks, D., Davis, A.M., Naglie, G., 2006. Validity of 3 physical performance measures in inpatient geriatric rehabilitation. Arch. Phys. Med. Rehabil. 87, 105–110. https://doi.org/10.1016/j.apmr.2005.08.109

Butland, R.J., Pang, J., Gross, E.R., Woodcock, A.A., Geddes, D.M., 1982. Two-, six-, and 12-minute walking tests in respiratory disease. Br. Med. J. (Clin. Res. Ed). 284, 1607–1608. https://doi.org/10.1136/bmj.284.6329.1607

Carter, R. E., Lubinsky, J., Domholdt, E., 2011. Rehabilitation research: Principles and applications. St. Louis, Mo. : Elsevier Saunders, c2011, St. Louis, Mo.

Chau, M.W.R., Chan, S.P., Wong, Y.W., Lau, M.Y.P., 2013. Reliability and validity of the Modified Functional Ambulation Classification in patients with hip fracture. Hong Kong Physiother. J. 31, 41–44. https://doi.org/10.1016/j.hkpj.2013.01.041

Chiu, H.F.K., Lee, H.C., Chung, W.S., Kwong, P.K., 1994. Reliability and validity of the Cantonese version of Mini-Mental State Examination - a preliminary study. Hong Kong J. Psychiatry 4, 25.

Collen, F.M., Wade, D.T., Bradshaw, C.M., 1990. Mobility after stroke: Reliability of measures of impairment and disability. Int. Disabil. Stud. 12, 6–9. https://doi.org/10.3109/03790799009166594

Connelly, D.M., Thomas, B.K., Cliffe, S.J., Perry, W.M., Smith, R.E., 2009. Clinical utility of the 2-minute walk test for older adults living in long-term care. Physiother. Canada 61, 78–87. https://doi.org/10.3138/physio.61.2.78

Crapo, R.O., Casaburi, R., Coates, A.L., Enright, P.L., MacIntyre, N.R., McKay, R.T., Johnson, D., Wanger, J.S., Zeballos, R.J., Bittner, V., Mottram, C., 2002. ATS statement: Guidelines for the six-minute walk test. Am. J. Respir. Crit. Care Med. 166, 111–117. https://doi.org/10.1164/rccm.166/1/111

Fitzpatrick, A.L., Buchanan, C.K., Nahin, R.L., Dekosky, S.T., Atkinson, H.H., Carlson, M.C., Williamson, J.D., 2007. Associations of gait speed and other measures of physical function with cognition in a healthy cohort of elderly persons. J. Gerontol. A. Biol. Sci. Med. Sci. 62, 1244–51. https://doi.org/10.1093/gerona/62.11.1244

Fitzpatrick, R., Davey, C., Buxton, M.J., Jones, D.R., 1998. Evaluating patient-based outcome measures for use in clinical trials. Health Technol. Assess. (Rockv). 2, i–iv, 1-74. https://doi.org/9812244

Flansbjer, U.B., Holmbäck, A.M., Downham, D., Patten, C., Lexell, J., 2005. Reliability of gait performance tests in men and women with hemiparesis after stroke. J. Rehabil. Med. 37, 75–82. https://doi.org/10.1080/16501970410017215

Fox, B., Henwood, T., Keogh, J., Neville, C., 2016. Psychometric viability of measures of functional performance commonly used for people with dementia. JBI Database Syst. Rev. Implement. Reports 14, 115–171. https://doi.org/10.11124/JBISRIR-2016-003064

Fox, B., Henwood, T., Neville, C., Keogh, J., 2014. Relative and absolute reliability of functional performance measures for adults with dementia living in residential aged care. Int. Psychogeriatrics 26, 1659–1667. https://doi.org/10.1017/S1041610214001124

Harada, N.D., Chiu, V., Stewart, A.L., 1999. Mobility-related function in older adults: Assessment with a 6-minute walk test. Arch. Phys. Med. Rehabil. 80, 837–841. https://doi.org/10.1016/S0003-9993(99)90236-8

Hollman, J.H., Beckman, B.A., Brandt, R.A., Merriwether, E.N., Williams, R.T., Nordrum, J.T., 2008. Minimum detectable change in gait velocity during acute rehabilitation following hip fracture. J. Geriatr. Phys. Ther. 31, 53–6.

Hoppes, S., Davis, L.A., Thompson, D., 2003. Environmental effects on the assessment of people with dementia: A pilot study. Am. J. Occup. Ther. 57, 396–402. https://doi.org/10.5014/ajot.57.4.396

Jabourian, A., Lancrenon, S., Delva, C., Perreve-Genet, A., Lablanchy, J.P., Jabourian, M., 2014. Gait velocity is an indicator of cognitive performance in healthy middle-aged adults. PLoS One 9. https://doi.org/10.1371/journal.pone.0103211

Kovach, C.R., Henschel, H., 1996. Planning Activities for Patients with Dementia: A Descriptive Study of Therapeutic Activities on Special Care Units. J. Gerontol. Nurs. 22, 33–38. https://doi.org/10.3928/0098-9134-19960901-10

Leung, S.O.C., Chan, C.C.H., Shah, S., 2007. Development of a Chinese version of the Modified Barthel Index — validity and reliability. Clin. Rehabil. 21, 912–922. https://doi.org/10.1177/0269215507077286

Maring, J.R., Costello, E., Birkmeier, M.C., Richards, M., Alexander, L.M., 2013. Validating functional measures of physical ability for aging people with intellectual developmental disability. Am. J. Intellect. Dev. Disabil. 118, 124–140. https://doi.org/10.1352/1944-7558-118.2.124

Miller, C. a, 2008. Communication difficulties in hospitalized older adults with dementia. Am. J. Nurs. 108, 58–66; quiz 67. https://doi.org/10.1097/01.NAJ.0000311828.13935.1e

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., de Vet, H.C.W., 2010. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J. Clin. Epidemiol. 63, 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Morley, J.E., Malmstrom, T.K., Miller, D.K., 2012. A simple frailty questionnaire (FRAIL) predicts outcomes in middle aged african americans. J. Nutr. Heal. Aging 16, 601–608. https://doi.org/10.1007/s12603-012-0084-2

Muir-Hunter, S.W., Graham, L., Odasso, M.M., 2015. Reliability of the berg balance scale as a clinical measure of balance in community-dwelling older adults with mild to moderate alzheimer disease: A pilot study. Physiother. Canada 67, 255–262. https://doi.org/10.3138/ptc.2014-32

Nordin, E., Rosendahl, E., Lundin-Olsson, L., 2006. Timed "Up & Go" test: reliability in older people dependent in activities of daily living--focus on cognitive state. Phys. Ther. 86, 646–55.

Perera, S., Mody, S.H., Woodman, R.C., Studenski, S.A., 2006. Meaningful change and responsiveness in common physical performance measures in older adults. J. Am. Geriatr. Soc. 54, 743–749. https://doi.org/10.1111/j.1532-5415.2006.00701.x

Peters, D.M., Fritz, S.L., Krotish, D.E., 2013. Assessing the reliability and validity of a shorter walk test compared with the 10-Meter Walk Test for measurements of gait speed in healthy, older adults. J. Geriatr. Phys. Ther. 36, 24–30. https://doi.org/10.1519/JPT.0b013e318248e20d

Pin, T.W., 2014. Psychometric Properties of 2-Minute Walk Test: A Systematic Review. Arch. Phys. Med. Rehabil. 95, 1759–1775. https://doi.org/10.1016/j.apmr.2014.03.034

Portney, L., Watkins, M., 2000. Foundations of clinical research: Applications to practice. Upper Saddle River, N.J. : Prentice Hall Health, c2000, Upper Saddle River, N.J.

Prosser, L., Canby, A., 1997. Further validation of EMS for measurement of mobility of hospitalised elderly people. Clin. Rehabil. 11, 338–43.

Ries, J.D., Echternach, J.L., Nof, L., Gagnon Blodgett, M., 2009. Test-retest reliability and minimal detectable change scores for the timed “up & go” test, the six-minute walk test, and gait speed in people with Alzheimer disease. Phys. Ther. 89, 569–79. https://doi.org/10.2522/ptj.20080258

Rikli, R.E., Jones, C.J., 1999. Development and Validation of a Functional Fitness Test for Community-Residing Older Adults. J. Aging Phys. Act. 7, 129–161. https://doi.org/10.1123/japa.7.2.129

Rockwood, K., Awalt, E., Carver, D., MacKnight, C., 2000. Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging. J. Gerontol. A. Biol. Sci. Med. Sci. 55, M70–M73. https://doi.org/10.1093/gerona/55.2.M70

Shah, S., Vanclay, F., Cooper, B., 1989. Improving the sensitivity of the Barthel Index for stroke rehabilitation. J. Clin. Epidemiol. 42, 703–709.

Shoukri, M.M., Asyali, M.H., Donner, A., 2004. Sample size requirements for the design of reliability study: review and new results. Stat. Methods Med. Res. 13, 251–271.

https://doi.org/10.1191/0962280204sm365ra

Small, J.A., Gutman, G., Makela, S., Hillhouse, B., 2003. Effectiveness of Communication Strategies Used by Caregivers of Persons With Alzheimer's Disease During Activities of Daily Living. J. Speech Lang. Hear. Res. 46, 353. https://doi.org/10.1044/1092-4388(2003/028)

Smith, R., 1994. Validation and reliability of the Elderly Mobility Scale. Physiotherapy 80, 744–747.

Steffen, T.M., Hacker, T. a, Mollinger, L., 2002. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. Phys. Ther. 82, 128–37. https://doi.org/10.1001/jama.1968.03140030033008

Tappen, R.M., Roach, K.E., Buchner, D., Barry, C., Edelstein, J., 1997. Reliability of physical performance measures in nursing home residents with Alzheimer's disease. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. 52, M52–M55.

Taylor, M.E., Delbaere, K., Lord, S.R., Mikolaizak, A.S., Brodaty, H., Close, J.C.T., 2014. Neuropsychological, physical, and functional mobility measures associated with falls in cognitively impaired older adults. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. 69, 987–995. https://doi.org/10.1093/gerona/glt166

Telenius, E.W., Engedal, K., Bergland, A., 2015. Inter-rater reliability of the Berg Balance Scale, 30 s chair stand test and 6 m walking test, and construct validity of the Berg Balance Scale in nursing home residents with mild-to-moderate dementia. BMJ Open 5, 1–7. https://doi.org/10.1136/bmjopen-2015-008321

Thomas, V.S., Hageman, P., 2002. A preliminary study on the reliability of physical performance measures in older day-care center clients with dementia. Int. Psychogeriatr. 14, 17–23. https://doi.org/10.1017/S1041610202008244

van Iersel, M.B., Benraad, C.E.M., Rikkert, M.G.M.O., 2007. Validity and reliability of quantitative gait analysis in geriatric patients with and without dementia. J. Am. Geriatr. Soc. 55, 632–634. https://doi.org/10.1111/j.1532-5415.2007.01130.x

van Iersel, M.B., Munneke, M., Esselink, R.A.J., Benraad, C.E.M., Olde Rikkert, M.G.M., 2008. Gait velocity and the Timed-Up-and-Go test were sensitive to changes in mobility in frail elderly patients. J. Clin. Epidemiol. 61, 186–191. https://doi.org/10.1016/j.jclinepi.2007.04.016

Watson, N.L., Rosano, C., Boudreau, R.M., Simonsick, E.M., Ferrucci, L., Sutton-Tyrrell, K., Hardy, S.E., Atkinson, H.H., Yaffe, K., Satterfield, S., Harris, T.B., Newman, A.B., 2010. Executive function, memory, and gait speed decline in well-functioning older adults. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. 65 A, 1093–1100. https://doi.org/10.1093/gerona/glq111

Weir, J.P., 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J. Strength Cond. Res. 19, 231–40. https://doi.org/10.1519/15184.1

Woo, J., Yu, R., Wong, M., Yeung, F., Wong, M., Lum, C., 2015. Frailty Screening in the Community Using the FRAIL Scale. J. Am. Med. Dir. Assoc. 16, 412–9. https://doi.org/10.1016/j.jamda.2015.01.087

Andresen, E.M., 2000. Criteria for assessing the tools of disability outcomes research. Arch. Phys. Med. Rehabil. 81, 15–20. https://doi.org/10.1053/apmr.2000.20619

Beck, C., Heacock, P., Rapp, C.G., Mercer, S.O., 1993. Assisting cognitively impaired elders with activities of daily living. Am. J. Alzheimers. Dis. Other Demen. 8, 11–20. https://doi.org/10.1177/153331759300800602

Beckerman, H., Roebroeck, M.E., Lankhorst, G.J., Becher, J.G., Bezemer, P.D., Verbeek, A.L.M., 2001. Smallest real difference, a link between reproducibility and responsiveness. Qual. Life Res. 10, 571–578. https://doi.org/10.1023/A:1013138911638

Berg, K.O., Wood-Dauphinee, S.L., Williams, J.I., Gayton, D., 1989. Measuring balance in elderly: preliminary development of an instrument. Physiother. Canada 41, 304–311.

Berg, K.O., Wood-Dauphinee, S.L., Williams, J.I., Maki, B., 1992. Measuring balance in the elderly: validation of an instrument. Can. J. Public Health 83 Suppl 2, S7-11.

Blankevoort, C.G., Scherder, E.J., Wieling, M.B., Hortobagyi, T., Brouwer, W.H., Geuze, R.H., van Heuvelen, M.J., 2013. Physical predictors of cognitive performance in healthy older adults: a cross-sectional analysis. PLoS One 8, e70799. https://doi.org/10.1371/journal.pone.0070799

Bohannon, R.W., Wang, Y.C., Gershon, R.C., 2015. Two-minute walk test performance by adults 18 to 85 years: Normative values, reliability, and responsiveness. Arch. Phys. Med. Rehabil. 96, 472–477. https://doi.org/10.1016/j.apmr.2014.10.006

Brooks, D., Davis, A.M., Naglie, G., 2006. Validity of 3 physical performance measures in inpatient geriatric rehabilitation. Arch. Phys. Med. Rehabil. 87, 105–110. https://doi.org/10.1016/j.apmr.2005.08.109

Butland, R.J., Pang, J., Gross, E.R., Woodcock, A.A., Geddes, D.M., 1982. Two-, six-, and 12-

minute walking tests in respiratory disease. Br. Med. J. (Clin. Res. Ed). 284, 1607–1608. https://doi.org/10.1136/bmj.284.6329.1607

Carter, R. E., Lubinsky, J., Domholdt, E., 2011. Rehabilitation research: Principles and applications. St. Louis, Mo. : Elsevier Saunders, c2011, St. Louis, Mo.

Chau, M.W.R., Chan, S.P., Wong, Y.W., Lau, M.Y.P., 2013. Reliability and validity of the Modified Functional Ambulation Classification in patients with hip fracture. Hong Kong Physiother. J. 31, 41–44. https://doi.org/10.1016/j.hkpj.2013.01.041

Chiu, H.F.K., Lee, H.C., Chung, W.S., Kwong, P.K., 1994. Reliability and validity of the Cantonese version of Mini-Mental State Examination - a preliminary study. Hong Kong J. Psychiatry 4, 25.

Collen, F.M., Wade, D.T., Bradshaw, C.M., 1990. Mobility after stroke: Reliability of measures of impairment and disability. Int. Disabil. Stud. 12, 6–9. https://doi.org/10.3109/03790799009166594

Connelly, D.M., Thomas, B.K., Cliffe, S.J., Perry, W.M., Smith, R.E., 2009. Clinical utility of the 2-minute walk test for older adults living in long-term care. Physiother. Canada 61, 78–87. https://doi.org/10.3138/physio.61.2.78

Crapo, R.O., Casaburi, R., Coates, A.L., Enright, P.L., MacIntyre, N.R., McKay, R.T., Johnson, D., Wanger, J.S., Zeballos, R.J., Bittner, V., Mottram, C., 2002. ATS statement: Guidelines for the six-minute walk test. Am. J. Respir. Crit. Care Med. 166, 111–117. https://doi.org/10.1164/rccm.166/1/111

Fitzpatrick, A.L., Buchanan, C.K., Nahin, R.L., Dekosky, S.T., Atkinson, H.H., Carlson, M.C., Williamson, J.D., 2007. Associations of gait speed and other measures of physical function with cognition in a healthy cohort of elderly persons. J. Gerontol. A. Biol. Sci. Med. Sci. 62, 1244–51. https://doi.org/10.1093/gerona/62.11.1244

Fitzpatrick, R., Davey, C., Buxton, M.J., Jones, D.R., 1998. Evaluating patient-based outcome measures for use in clinical trials. Health Technol. Assess. (Rockv). 2, i–iv, 1-74. https://doi.org/9812244

Flansbjer, U.B., Holmbäck, A.M., Downham, D., Patten, C., Lexell, J., 2005. Reliability of gait performance tests in men and women with hemiparesis after stroke. J. Rehabil. Med. 37, 75–82. https://doi.org/10.1080/16501970410017215

Fox, B., Henwood, T., Keogh, J., Neville, C., 2016. Psychometric viability of measures of functional performance commonly used for people with dementia. JBI Database Syst. Rev. Implement. Reports 14, 115–171. https://doi.org/10.11124/JBISRIR-2016-003064

Fox, B., Henwood, T., Neville, C., Keogh, J., 2014. Relative and absolute reliability of functional performance measures for adults with dementia living in residential aged care. Int. Psychogeriatrics 26, 1659–1667. https://doi.org/10.1017/S1041610214001124

Harada, N.D., Chiu, V., Stewart, A.L., 1999. Mobility-related function in older adults: Assessment with a 6-minute walk test. Arch. Phys. Med. Rehabil. 80, 837–841. https://doi.org/10.1016/S0003-9993(99)90236-8

Hollman, J.H., Beckman, B.A., Brandt, R.A., Merriwether, E.N., Williams, R.T., Nordrum, J.T., 2008. Minimum detectable change in gait velocity during acute rehabilitation following hip fracture. J. Geriatr. Phys. Ther. 31, 53–6.

Hoppes, S., Davis, L.A., Thompson, D., 2003. Environmental effects on the assessment of people with dementia: A pilot study. Am. J. Occup. Ther. 57, 396–402. https://doi.org/10.5014/ajot.57.4.396

Jabourian, A., Lancrenon, S., Delva, C., Perreve-Genet, A., Lablanchy, J.P., Jabourian, M., 2014. Gait velocity is an indicator of cognitive performance in healthy middle-aged adults. PLoS One 9. https://doi.org/10.1371/journal.pone.0103211

Kovach, C.R., Henschel, H., 1996. Planning Activities for Patients with Dementia: A Descriptive Study of Therapeutic Activities on Special Care Units. J. Gerontol. Nurs. 22, 33–38. https://doi.org/10.3928/0098-9134-19960901-10

Leung, S.O.C., Chan, C.C.H., Shah, S., 2007. Development of a Chinese version of the Modified Barthel Index — validity and reliability. Clin. Rehabil. 21, 912–922. https://doi.org/10.1177/0269215507077286

Maring, J.R., Costello, E., Birkmeier, M.C., Richards, M., Alexander, L.M., 2013. Validating functional measures of physical ability for aging people with intellectual developmental disability. Am. J. Intellect. Dev. Disabil. 118, 124–140. https://doi.org/10.1352/1944-7558-118.2.124

Miller, C. a, 2008. Communication difficulties in hospitalized older adults with dementia. Am. J. Nurs. 108, 58–66; quiz 67. https://doi.org/10.1097/01.NAJ.0000311828.13935.1e

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., de Vet, H.C.W., 2010. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J. Clin. Epidemiol. 63, 737–745.

https://doi.org/10.1016/j.jclinepi.2010.02.006

Morley, J.E., Malmstrom, T.K., Miller, D.K., 2012. A simple frailty questionnaire (FRAIL) predicts outcomes in middle aged african americans. J. Nutr. Heal. Aging 16, 601–608. https://doi.org/10.1007/s12603-012-0084-2

Muir-Hunter, S.W., Graham, L., Odasso, M.M., 2015. Reliability of the berg balance scale as a clinical measure of balance in community-dwelling older adults with mild to moderate alzheimer disease: A pilot study. Physiother. Canada 67, 255–262. https://doi.org/10.3138/ptc.2014-32

Nordin, E., Rosendahl, E., Lundin-Olsson, L., 2006. Timed “Up & Go” test: reliability in older people dependent in activities of daily living--focus on cognitive state. Phys. Ther. 86, 646–55.

Perera, S., Mody, S.H., Woodman, R.C., Studenski, S.A., 2006. Meaningful change and responsiveness in common physical performance measures in older adults. J. Am. Geriatr. Soc. 54, 743–749. https://doi.org/10.1111/j.1532-5415.2006.00701.x

Peters, D.M., Fritz, S.L., Krotish, D.E., 2013. Assessing the reliability and validity of a shorter walk test compared with the 10-Meter Walk Test for measurements of gait speed in healthy, older adults. J. Geriatr. Phys. Ther. 36, 24–30. https://doi.org/10.1519/JPT.0b013e318248e20d

Pin, T.W., 2014. Psychometric Properties of 2-Minute Walk Test: A Systematic Review. Arch. Phys. Med. Rehabil. 95, 1759–1775. https://doi.org/10.1016/j.apmr.2014.03.034

Portney, L., Watkins, M., 2000. Foundations of clinical research: Applications to practice. Upper

Saddle River, N.J. : Prentice Hall Health, c2000, Upper Saddle River, N.J.

Prosser, L., Canby, A., 1997. Further validation of EMS for measurement of mobility of hospitalised elderly people. Clin. Rehabil. 11, 338–43.

Ries, J.D., Echternach, J.L., Nof, L., Gagnon Blodgett, M., 2009. Test-retest reliability and minimal detectable change scores for the timed “up & go” test, the six-minute walk test, and gait speed in people with Alzheimer disease. Phys. Ther. 89, 569–79. https://doi.org/10.2522/ptj.20080258

Rikli, R.E., Jones, C.J., 1999. Development and Validation of a Functional Fitness Test for Community-Residing Older Adults. J. Aging Phys. Act. 7, 129–161. https://doi.org/10.1123/japa.7.2.129

Rockwood, K., Awalt, E., Carver, D., MacKnight, C., 2000. Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging. J. Gerontol. A. Biol. Sci. Med. Sci. 55, M70–M73. https://doi.org/10.1093/gerona/55.2.M70

Shah, S., Vanclay, F., Cooper, B., 1989. Improving the sensitivity of the Barthel Index for stroke rehabilitation. J. Clin. Epidemiol. 42, 703–709.

Shoukri, M.M., Asyali, M.H., Donner, A., 2004. Sample size requirements for the design of reliability study: review and new results. Stat. Methods Med. Res. 13, 251–271. https://doi.org/10.1191/0962280204sm365ra

Small, J.A., Gutman, G., Makela, S., Hillhouse, B., 2003. Effectiveness of Communication Strategies Used by Caregivers of Persons With Alzheimer’s Disease During Activities of

Daily Living. J. Speech Lang. Hear. Res. 46, 353. https://doi.org/10.1044/1092-4388(2003/028)

Smith, R., 1994. Validation and reliability of the Elderly Mobility Scale. Physiotherapy 80, 744–747.

Steffen, T.M., Hacker, T. a, Mollinger, L., 2002. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. Phys. Ther. 82, 128–37. https://doi.org/10.1001/jama.1968.03140030033008

Tappen, R.M., Roach, K.E., Buchner, D., Barry, C., Edelstein, J., 1997. Reliability of physical performance measures in nursing home residents with Alzheimer's disease. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. 52, M52–M55.

Taylor, M.E., Delbaere, K., Lord, S.R., Mikolaizak, A.S., Brodaty, H., Close, J.C.T., 2014. Neuropsychological, physical, and functional mobility measures associated with falls in cognitively impaired older adults. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. 69, 987–995. https://doi.org/10.1093/gerona/glt166

Telenius, E.W., Engedal, K., Bergland, A., 2015. Inter-rater reliability of the Berg Balance Scale, 30 s chair stand test and 6 m walking test, and construct validity of the Berg Balance Scale in nursing home residents with mild-to-moderate dementia. BMJ Open 5, 1–7. https://doi.org/10.1136/bmjopen-2015-008321

Thomas, V.S., Hageman, P., 2002. A preliminary study on the reliability of physical performance measures in older day-care center clients with dementia. Int. Psychogeriatr. 14, 17–23. https://doi.org/10.1017/S1041610202008244 van Iersel, M.B., Benraad, C.E.M., Rikkert, M.G.M.O., 2007. Validity and reliability of quantitative gait analysis in geriatric patients with and without dementia. J. Am. Geriatr. Soc. 55, 632–634. https://doi.org/10.1111/j.1532-5415.2007.01130.x

van Iersel, M.B., Munneke, M., Esselink, R.A.J., Benraad, C.E.M., Olde Rikkert, M.G.M., 2008. Gait velocity and the Timed-Up-and-Go test were sensitive to changes in mobility in frail elderly patients. J. Clin. Epidemiol. 61, 186–191. https://doi.org/10.1016/j.jclinepi.2007.04.016

Watson, N.L., Rosano, C., Boudreau, R.M., Simonsick, E.M., Ferrucci, L., Sutton-Tyrrell, K., Hardy, S.E., Atkinson, H.H., Yaffe, K., Satterfield, S., Harris, T.B., Newman, A.B., 2010. Executive function, memory, and gait speed decline in well-functioning older adults. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. 65 A, 1093–1100. https://doi.org/10.1093/gerona/glq111

Weir, J.P., 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J. Strength Cond. Res. 19, 231–40. https://doi.org/10.1519/15184.1

Woo, J., Yu, R., Wong, M., Yeung, F., Wong, M., Lum, C., 2015. Frailty Screening in the Community Using the FRAIL Scale. J. Am. Med. Dir. Assoc. 16, 412–9. https://doi.org/10.1016/j.jamda.2015.01.087

Table 1. Characteristics of participants

Age (years)	
Mean (SD)	87.1 ± 6.2
Gender, n (%)	
Male	3 (7.7)
Female	36 (92.3)
Setting, n (%)	
Day care center	19 (48.7)
Residential care facility	20 (51.3)
Body mass index (kg/m^2)	
Mean (SD)	22.0 ± 3.3
Number of chronic diseases	
Mean (SD)	5.9 ± 2.9
MMSE (0-30)*	
Mean (SD)	13.2 ± 5.5
Use of walking aids, n (%)	
Unaided	17 (43.6)
Stick	9 (23.1)
Quadripod	3 (7.7)
Rollator	9 (23.1)
Frame	1 (2.6)

Ambulatory status, n (%)	
Indoor walker	11 (28.2)
Outdoor walker	28 (71.8)
EMS (0-20)*	
Mean (SD)	17.4 ± 3.0
BBS (0-56)*	
Mean (SD)	44.1 ± 10.0
MBI (0-100)*	
Mean (SD)	86.5 ± 9.1

Notes: * The bracket indicates the possible range of score of the outcome measures. A higher score reflects better outcome.

EMS- Elderly Mobility Scale; BBS- Berg Balance Scale; MBI- Modified Barthel Index; MMSE- Mini-Mental State Examination

907

Table 2. Mean (SD), test-retest and inter-rater reliability coefficients of the walk tests

	Mean (SD)			Test-rest reliability coefficient	Inter-rater reliability coefficient	
				$ICC_{2,1}$ (95% CI)	$ICC_{3,2}$ (95% CI)	
	Assessor A Occasion 1	Assessor A Occasion 2	Assessor B		Assessor A Occasion 1 and Assessor B	Assessor A Occasion 2 and Assessor B
2MWT (m)	63.3 ± 21.6	63.1 ± 22.0	63.6 ± 22.8	.98 (.96-.99)	.92 (.86-.96)	.96 (.92-.98)
6MWT (m)	194.0 ± 79.4	198.6 ± 77.0	203.6 ± 80.0	.98 (.97-.99)	.95 (.91-.97)	.94 (.89-.97)
10MeWT-2M (m/s)	.63 ± .21	.60 ± .21	.62 ± .23	.91 (.83-.95)	.86 (.75-.93)	.93 (.87-.96)
10MeWT-6M (m/s)	.64 ± .26	.65 ± .22	.73 ± .29	.94 (.89-.97)	.65 (.42-.80)	.60 (.36-.77)

Notes: 2MWT- 2-minute walk test; 6MWT- 6-minute walk test; 10MeWT-2M- 10-meter walk test (measured in 2-minute walk test); 10MeWT-6M- 10-meter walk test (measured in 6-minute walk test)

Table 3. Correlations between the walk tests and functional measures

	2MWT	6MWT	10MeWT-2M	10MeWT-6M
2MWT				
ρ	1.00			
p-value				
6MWT				
ρ	.93	1.00		
p-value	<.001			
10MeWT-2M				
ρ	.91	.87	1.00	
p-value	<.001	<.001		
10MeWT-6M				
ρ	.84	.91	.84	1.00
p-value	<.001	<.001	<.001	
EMS				
ρ	.43	.39	.39	.27
p-value	.007	.014	.014	.104

BBS				
ρ	.49	.47	.49	.35
p-value	.002	.002	.002	.029
MBI				
ρ	.54	.48	.45	.46
p-value	<.001	.002	.004	.004

Notes: The results of the walk tests done by Assessor A in occasion 1 were used for the analyses.

*Spearman *ρ* correlation

2MWT- 2-minute walk test; 6MWT- 6-minute walk test; 10MeWT-2M- 10-meter walk test (measured in 2-minute walk test); 10MeWT-6M- 10-meter walk test (measured in 6-minute walk test) EMS- Elderly Mobility Scale; BBS- Berg Balance Scale; MBI- Modified Barthel Index

Table 4. Comparisons between subgroups of the participants

			2MWT (m)	6MWT (m)	10MeWT-2M (m/s)	10MeWT-6M (m/s)
Walking aids, mean (SD)						
No aid (n=17)			75.8 ± 19.1	241.0 ± 71.2	.72 ± .19	.75 ± .26
Stick (n=9)			53.6 ± 12.4	168.6 ± 40.3	.57 ± .15	.53 ± .13
Quadripod or frame (n=13)			53.7 ± 22.5	150.2 ± 80.2	.54 ± .25	.58 ± .28
*p-value**	Overall		.004	.002	.047	.060
	Pairwise	No aid vs stick	.023	.046	.24	.10
		No aid vs quadripod or frame	.010	.003	.065	.20

		Stick vs quadripod or frame	1.00	1.00	1.00	1.00
Ambulatory status, mean (SD)						
Outdoor walker (n=11)			78.6 ± 21.6	252.5 ± 79.8	.74 ± .23	.80 ± .28
Indoor walker (n=28)			57.3 ± 18.8	171.0 ± 67.7	.58 ± .20	.58 ± .22
*p-value**			.004	.003	.036	.017

Notes: The results of the walk tests done by Assessor A in occasion 1 were used.

*One-way ANOVA with Bonferroni adjustment or independent t-test were used.

2MWT- 2-minute walk test; 6MWT- 6-minute walk test; 10MeWT-2M- 10-meter walk test (measured in 2-minute walk test); 10MeWT-6M- 10-meter walk test (measured in 6-minute walk test)

Table 5. SEM, MDC_{95} and MDC_{95}% of the walk tests

Variables	SEM	MDC_{95}	MDC_{95}% (%)
2MWT (m)	3.3	9.1	14.4
6MWT (m)	10.1	28.1	14.3
10MeWT-2M (m/s)	.06	.17	28.5
10MeWT-6M (m/s)	.06	.16	24.4

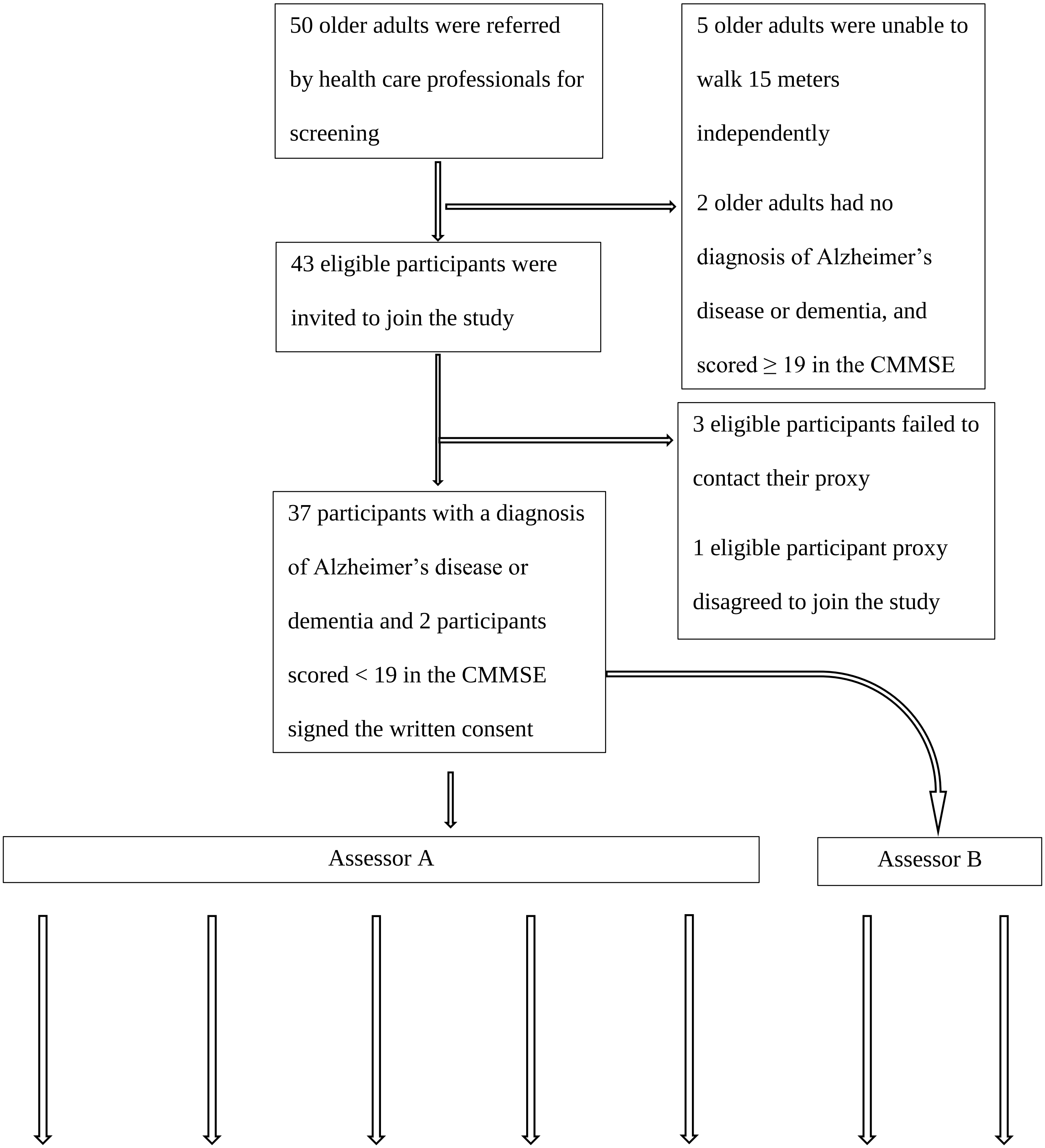
Notes: 2MWT- 2-minute walk test; 6MWT- 6-minute walk test; 10MeWT-2M- 10-meter walk test (measured in 2-minute walk test); 10MeWT-6M- 10-meter walk test (measured in 6-minute walk test); SEM- standard error of measurement; MDC_{95}- minimal detectable change at 95% confidence interval

Table 6. Median (interquartile range), test-retest and inter-rater reliability coefficients of the cues provided by the assessors in the walk tests

	Median (interquartile range)			Test-rest reliability coefficient	Inter-rater reliability coefficient	
				$ICC_{2,1}$ (95% CI)	$ICC_{3,2}$ (95% CI)	
	Assessor A Occasion 1	Assessor A Occasion 2	Assessor B		Assessor A Occasion 1 and Assessor B	Assessor A Occasion 2 and Assessor B
2MWT and 10MeWT-2M	1.00 (2.00)	1.00 (2.00)	1.00 (1.00)	.88 (.78-.93)	.83 (.70-.91)	.62 (.38-.78)
6MWT and 10MeWT-6M	1.00 (2.00)	1.00 (3.00)	1.00 (2.00)	.89 (.80-.94)	.83 (.69-.91)	.69 (.47-.82)

Notes: 2MWT- 2-minute walk test; 6MWT- 6-minute walk test; 10MeWT-2M- 10-meter walk test (measured in 2-minute walk test); 10MeWT-6M- 10-meter walk test (measured in 6-minute walk test)

Figure 1. Participant recruitment process and study procedure



Occasion 1*‡^	Occasion 2*	Occasion 3†#^	Occasion 4†	Occasion 5^	Occasion 6‡	Occasion 7#
2MWT & 10MeWT-2M	2MWT & 10MeWT-2M	6MWT & 10MeWT-6M	6MWT & 10MeWT-6M	BBS, EMS and MBI	2MWT & 10MeWT-2M	6MWT & 10MeWT-6M

*Data collected from Occasion 1 and 2 were used to analyze the test-retest reliability of the 2MWT and 10MeWT-2M

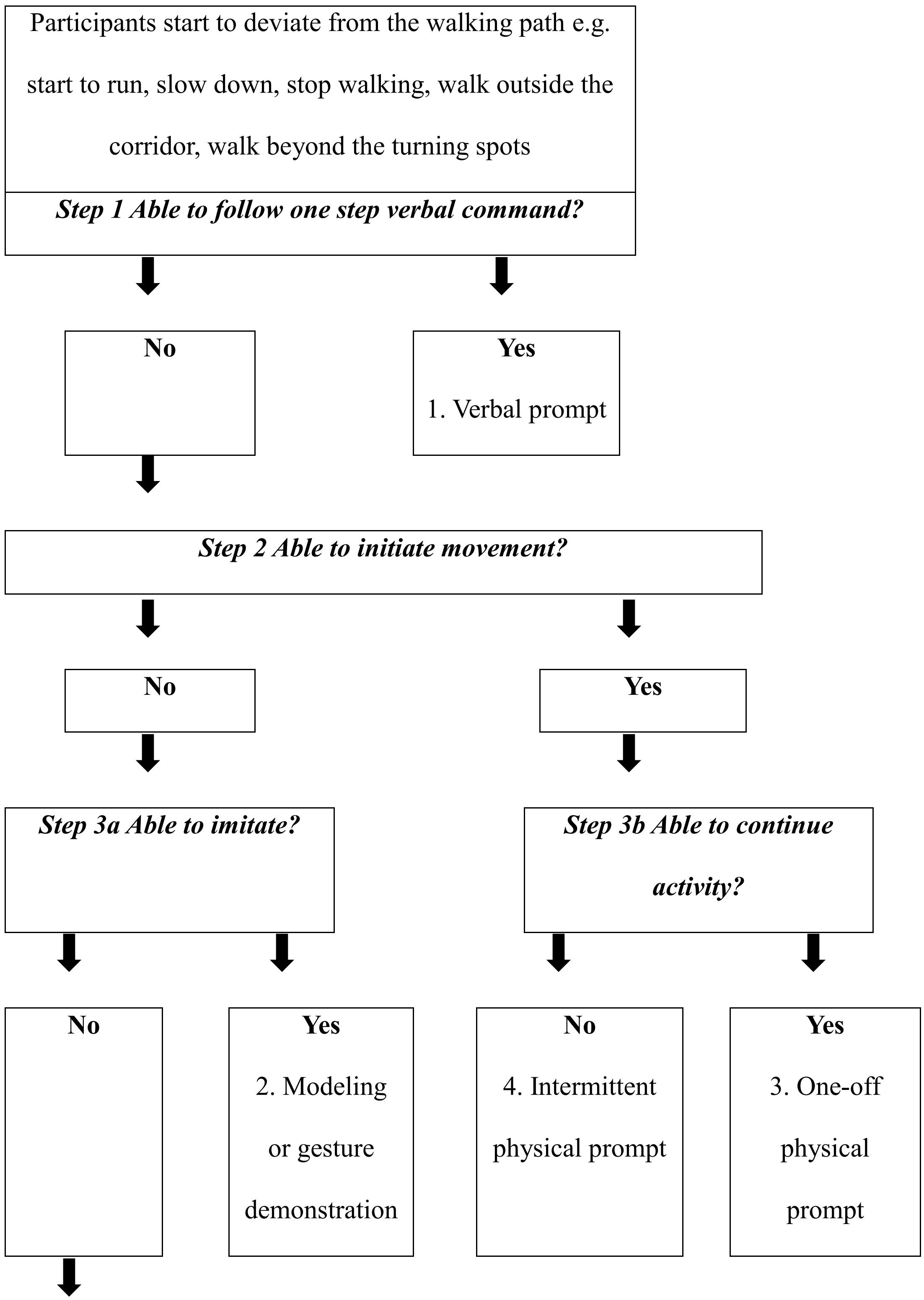
†Data collected from Occasion 3 and 4 were used to analyze the test-retest reliability of the 6MWT and 10MeWT-6M

‡Data collected from Occasion 1 and 6 were used to analyze the inter-rater reliability of the 2MWT and 10MeWT-2M

#Data collected from Occasion 3 and 7 were used to analyze the inter-rater reliability of the 6MWT and 10MeWT-6M

^Data collected from Occasion 1, 3 and 5 were used to analyze the construct validity of all the walk tests

Figure 2. The progressive cueing system in the walk tests



Step 4 Able to continue activity?

No	**Yes**
6. Complete physical guidance	5. Intermittent physical guidance