Geriatric Nursing 55 (2024) 71–78

Contents lists available at ScienceDirect

Geriatric Nursing

journal homepage: www.gnjournal.com

# Comparing remote and face-to-face assessments of physical performance in older adults: A reliability study

Fanny Buckinx, PhD[a,*], Marvin Rezoulat, MSc[b], César Lefranc, MSc[b], Jean-Yves Reginster, MDPhD[a], Olivier Bruyere, PhD[a,b]

[a] WHO Collaborating Center for Epidemiologic aspects of musculo-skeletal health and ageing, Division of Public Health, Epidemiology and Health Economics, University of Liège, Belgium

[b] Department of physical activity and rehabilitation sciences, University of Liège, Belgium

A R T I C L E I N F O

*Article history:*
Received 9 October 2023
Received in revised form 3 November 2023
Accepted 7 November 2023
Available online xxx

*Keywords:*
Physical performance
Online assessment
Older people
Reliability
Validity

A B S T R A C T

*Introduction:* Older people often experience a decline in their physical performance. Tests have been approved to evaluate this performance in person. Yet, the constraints associated with in-person assessments (e.g. lack of medical facilities, pandemic lockdown, and contagion risk) are making us contemplate setting up assessments remotely.

*Objectives:* To determine whether remote physical performance measurements of older adults are reliable and valid compared to face-to-face measurements.

*Methods:* Forty-five subjects aged 65 and over completed the normal/fast speed test (NWT/FWT), the unipodal balance test (UBT), the normal/fast timed up and go test (NTUG/FTUG), the 5 and 10 rep sit to stand test (5STS and 10STS), the 30 sec chair stand (30CS), the 2 minute step test (2MST) and the flexibility before standing (SAD) once face-to-face and twice remotely, by two different observers. The intraclass correlation coefficients (ICC), the standard errors of measurement (SEM%) and minimum detectable changes (MDC%) were calculated for both intra- and inter-observer conditions, to assess the relative and the absolute reliability. An ICC value exceeding 0.90 indicates a very high reliability, while an ICC between 0.70 and 0.89 signifies a high reliability. In clinical practice, a SEM % of less than 10% is considered acceptable. A smaller MDC % indicates a measurement that is more sensitive to detecting changes.

*Results:* Intra-observer relative reliability was very high (ICC>0.9) for the UBT, NWT, NTUG, FTUG, 5STS, 10STS, 30CS and the SAD; and high (ICC>0.7) for the 2MST and FWS. SEM% values ranged from 0% to 24.03% and MDC from 0% to 9.93%. Inter-observer relative reliability was considered very high (ICC>0.9) for all tests. SEM% values ranged from 0% to 17.68% and MDC from 0% to 7.32%.

*Conclusion:* Our findings demonstrate that remote assessments exhibited consistently high to very high levels of intra- and inter-observer relative reliability when compared to face-to-face assessments. Additionally, certain remote evaluations showed acceptable absolute reliability, making them viable alternatives for healthcare professionals when in-person assessments are not feasible in clinical practice.



## Introduction

Traditionally, face-to-face assessments have been the standard approach to assessing the physical function in older adults.[1] Physical function refers to an individual's capacity to perform various physical tasks, which can span from basic self-care activities (such as daily living tasks) to more intricate activities that demand a combination of skills, often involving social interactions or taking place within a social context.[2] Typically, physical function is assessed objectively through physical performance tests.[3] These tests are essential for detecting impairments, preventing further deterioration in physical health, and improving the overall quality of life of older people.[4,5] They also help healthcare professionals to make informed decisions about the appropriate level of care and support that older adults need to maintain their independence and well-being.[6] These face-to-face evaluations have proven to be valuable in accurately assessing functional capacity and guiding appropriate interventions.[1] However, challenges such as limited access to healthcare facilities, geographical limitations, and social isolation, can prevent older adults from participating in these face-to-face assessments. Furthermore, the emergence of the COVID-19 pandemic posed unprecedented challenges to

*Corresponding author at: WHO Collaborating Center for Public Health aspects of musculo-skeletal health and ageing, Division of Public Health, Epidemiology and Health Economics, University of Liège, Avenue Hippocrate 13, CHU Bât B23, 4000 Liège, Belgium

*E-mail address:* fanny.buckinx@uliege.be (F. Buckinx).

https://doi.org/10.1016/j.gerinurse.2023.11.004
0197-4572/$ – see front matter 

healthcare systems worldwide, requiring stringent safety measures, including social and physical distancing, which further limited face-to-face interactions.[7,8]

Amidst these challenges, the evolution of technology has opened new avenues for conducting remote assessments of physical performance in older adults.[9] Remote assessments, facilitated by web-based technologies, has the potential to overcome the barriers faced by this vulnerable population, by providing more accessible and convenient assessment options.[9,10] The use of telehealth and virtual assessment has gained momentum in recent years and has been shown to be effective and well accepted in various healthcare settings, including geriatric care.[11,12]

Nonetheless, the absolute (i.e. reliability of scores within individual participants on different occasions[13]) and relative (i.e. the linearity of the relationship between two repeated measures[13]) reliability of remote evaluations compared to face-to-face assessments require careful consideration. While the feasibility of remote assessment has been demonstrated in various populations,[14-20] the specific context of physical performance assessment in older adults remains understudied. A recent Canadian study suggested that assessing physical performance remotely in older adults seems feasible and reliable compared with face-to-face assessments although future studies are needed to confirm this conclusion.[14] In fact, only 15 subjects were included in this study and a limited number of physical tests were evaluated.[14] Therefore, rigorous investigation is needed to ensure that remote assessments provide accurate and reliable results to enable informed clinical decision making.

In this context, our study aimed to investigate whether remote assessment could provide comparable results to face-to-face assessment in measuring the physical performance in older adults. In this study, 10 tests were evaluated (the normal/fast speed test (NWT/FWT), the unipodal balance test (UBT), the normal/fast timed up and go test (NTUG/FTUG), the 5 and 10 rep sit to stand test (5STS and 10STS), the 30 sec chair stand (30CS), the 2 minute step test (2MST), and the flexibility before standing (SAD) test. We hypothesized that remote assessment of physical performance would show similar reliability to face-to-face assessment, particularly for walking speed, which was our main criterion (statistical power was calculated on this criterion).

By investigating the comparability of remote and face-to-face assessments of physical performance in older adults, our research aims to provide valuable insights for clinicians, researchers, and policy makers in shaping future geriatric care strategies.

## Methods

### *Study design*

A cross-sectional tool validation study was conducted. Data collection took place from January 2023 to April 2023. The study was approved by the Ethics Committee of the University Hospital of Liège (B7072021000085) and all participants signed an informed consent form before to the study.

### *Population*

Volunteer community-dwelling older people were recruited to participate in this study through social media. To be included in this study, participants had to (a) have an email address, an internet connection, and a digital device with a webcam at home (i.e. computer, laptop, tablet or smartphone), (b) be aged 65 years and older, (c) live independently in the community (Barthel score = 100),[21] (d) be able to perform a physical performance test, (e) be able to give consent (no cognitive impairment based on the Mini Mental State Examination : MMSE $\geq$ 24 [22]).

The normal walking speed test was used as the main criterion of physical performance to compare the results obtained by the "online" assessment to those obtained by the "face-to-face" assessment. Thus, based on an average walking speed of 0.98 $\pm$ 0.29 seconds (based on the results of the SarcoPhage cohort, a Belgian cohort of community dwelling older adults[23]) for our population and by tolerating a maximum difference (data defined arbitrarily) of 0.2 seconds in the walking speed test between the 2 evaluation methods (online vs. face-to-face), 45 subjects were necessary to successfully complete this first step of the research project (with a statistical power of 90 % and a level of certainty $\alpha$ = 0.05).

### *Measures*

#### *Socio-demographic characteristics*

The socio-demographic characteristics of the participants were self-reported during a telephone call prior to the physical performance assessments: age, gender, weight and height (to calculate body mass index (BMI)), country of residence and Socio-Economic Position Index (= age - 6 x education level - 4 x Professional Category + 55).[24]

#### *Physical performances*

The 10 tests described in Table 1 have been validated in the present study and have been selected from test batteries regularly used to assess physical performance in older people.

### ***Evaluation protocols***

These physical performance tests were performed at baseline and after 3 days by the first observer and after 6 days by the second observer. The selection of a three-day period for re-assessment was based on several considerations, including the need to minimize potential day-to-day variations in participants' performance and to ensure a representative evaluation of intra-rater reliability over a short time frame. Additionally, this choice was influenced by a similar study conducted by Peyrusqué et al. in Canada, which also used a three-day interval for re-assessment.[14] Thus, the test–retest for intra- (baseline versus day 3) and inter- (day 3 versus day 6) observers was performed to assess the reliability of the tests (Figure 1).

Remote assessments were conducted using Teams® via a computer, a laptop a tablet or a mobile phone while face-to-face assessments were conducted at the participant's home. Each assessment lasted approximately one hour. For safety reasons, participants were asked to wear closed shoes with grippy soles. The floor had to be non-slip and potential obstacles had to be kept out of the way. Participants were also asked to stand close to a support (wall, table). Finally, participants were asked to have a glass of water and a chair to sit on. For the NWS and TUG, participants were instructed to measure the distances (4 m and 3 m respectively) and make marks on the ground at the measured locations. For the 2MST, the height of the iliac crest and the patella were measured and marked on a wall. A piece of tape was then stuck to the wall at half the distance between the two. Each participant was instructed to step up and down and raise each knee to the mark on the wall as many times as possible during the 2-minute period. Regarding the positioning and configuration of the digital tool, participants were asked to place it on the floor or on a raised support. The video was taken from all angles to visualise the whole body and any markings. The tool was placed in front of the subject or on its side, with full visualisation of the movement being performed. Finally, the sound was adjusted to suit the participants.

### *Statistical analysis*

A Shapiro–Wilk test was used to verify the normal distribution for all parameters. Quantitative variables were expressed as

**Table 1**
Physical performances tests performed in this study

| Name of the test | Evaluated parameter | Description of the test | Interpretation | Metrological properties from previously published studies |
|---|---|---|---|---|
| Normal Walking Speed [40-42] | Walking speed | 1. Stand before the start line.<br>2. Cross the start line → *START stopwatch*<br>3. Walk at normal pace for 4m<br>4. *STOP stopwatch* when crossing the finish line | • Score (m/s):<br>- 66-70 years : W = 0.70 / M = 0.73<br>- > 70 years : W = 0.65 / M = 0.72 | • Test-retest reliability : ICC = 0.94<br>• SEM : 2 sec<br>• Sensitivity : N/A<br>• Specificity : N/A<br>• MDC : N/A |
| Fast Walking Speed [40-42] | Walking speed | Same as for Normal Walking Speed but at fast pace | Same as for Normal Walking Speed | N/A |
| Unipedal balance test [43-45] | Unipedal balance | 1. Stand with arms at sides<br>2. Remove one foot → *START stopwatch*<br>3. Maintain one-foot position<br>4. *STOP stopwatch* when: foot touches the ground, legs touch each other, arms become unbalanced | • Time < 5 seconds : risk of fall<br>• Normative data (seconds) :<br>- 60-69 years : W = 30.4 / M = 33.8<br>- 70-79 years : W = 16 .7 / M = 25.9 80-99 years : W = 10.6 / M = 8.7 | • Inter-observer reliability : ICC= 0.99<br>• SEM : 8.7 sec (15)<br>• SEM (%) : 40.8 (15)<br>• Sensitivity : 91%<br>• Specificity : 75% (16)<br>• MDC : 24.1<br>• MDC (%) : 113.1(15) |
| Normal Timed Up and Go test [30, 46-48] | Walking parameters | 1. Sitting on a chair<br>2. Raise chair → *START stopwatch*<br>3. Walk 3m<br>4. U-turn<br>5. Return to the chair and sit down → *STOP stopwatch* | • Time (s) :<br>∘ 60-69 years : 8.1<br>∘ 70-79 years : 9.2<br>∘ 80-99 years : 11.3<br>• Time :<br>- < 10 sec : total independance<br>- < 20 sec : independance for the main transfers<br>- < 30 sec : need assitance<br>• Time> 13.5 sec : Risk of fall | • Intra- and inter- observers reliability : ICC = 0.99<br>• SEM : 0.5<br>• Sensitivity : 87%<br>• Specificity : 87%<br>• MDC : N/A |
| Fast Timed Up and Go test [30, 46-48] | Walking parameters | Same as for Normal Timed Up and Go test but at fast pace | Same as for normal Timed Up and Go test | N/A |
| 5 Rep Sit to Stand [49-51] | Lower limbs functional capacity | 1. Sitting on a chair, arms crossed over your chest<br>2. Stand up (→ START stopwatch) and sit down 5 times in a row<br>3. Stop the stopwatch → On the 5th rise (standing position) | • Time (sec) :<br>- 60-69 years : 11.4<br>- 70-79 years : 12.6<br>- 80-89 years : 14.8<br>• 12 sec → Need to assess the rik of fall<br>• 15 sec → Risk of recurrent fallsN | • Intra-observer reliability: ICC= 0.91-0.933<br>• Inter-observers reliability : ICC= 0.99<br>• SEM : N/A<br>• Sensitivity : 55%<br>• Specificity : 65%<br>• MDC : N/A |
| 10 Rep Sit to Stand [52] | Lower limb muscle strength | Same as for the 5 Rep Sit To Stand BUT STOP the stopwatch → On the 10th lift (standing position) | • Power (W) – Mean value according to Takai et al: 184 | • Reliability : ICC= 0.88-0.96<br>• SEM : N/A<br>• Sensitivity : N/A<br>• Specificity : N/A<br>• MDC : N/A |
| 30 sec Chair Stand [31, 52-54] | Lower limbs endurance | Same as for the 5 Rep Sit To Stand BUT STOP the stopwatch → After 30 seconds | • Number of repetitions:<br>- 60-64 years: M = <14 / W = <12<br>- 65-69ans : M = <12 / W = <11<br>- 70-74ans : H = <12 / F = <10<br>- 75-79ans : H = <11 / F = <10<br>- 80-84ans = H = <10 / F = <9<br>- 85-89ans : H = <8 / F = <8<br>- 90-94ans : H = <7 / F = <4 | • Reliability (ICC) : 0.94<br>• SEM : N/A<br>• Sensitivity : 76.4% (F) / 75% (H)<br>• Specificity : 76.8% (F) / 71.7 % (H)<br>• MDC : 3.3 sec |
| 2 minutes Step Test [55, 56] | Global endurance | Placement of a marker at a height = mid-distance between femoral head and knee.<br>1. Patient standing with arms at sides next to a wall.<br>2. Alternate knee raises (→ START stopwatch) (left and right) at the marker.<br>3. Stop the stopwatch → After 2 minutes | • Number of repetitions:<br>- 60-64 years : M= 87-115 / W = 75-107<br>- 65-79 years : M= 86-116 / W = 73-107<br>- 70-74 years : M = 80-110 / W = 68-101<br>- 75-79 years : M = 73-109 / W = 68-100<br>- 80-84 years : M = 71-103 / W = 60-90<br>- 85-90 years : M = 59-91 / W = 55-85<br>- 90-95 years : M = 52-86 / W = 44-72 | • Reliability (ICC) : 0.90<br>• SEM : N/A<br>• Sensitivity : 92.24%<br>• Specificity : 81.36%<br>• MDC : N/A |
| Stand and Reach flexibility test [57] | Flexibility | 1. Trunk bent forward, knees straight<br>2. Arms along the legs, touching as low as possible<br>3. STOP measurement → at the distance limit reached | • Level reached : - Ground with fingers closed: M=5 / W=4- Fingertips touch the ground: M=4 / W=3- Fingertips reach bottom of shins: M=2 / W= 1- Fingertips reach instep: W=2- Fingertips reach middle of shins: M=1 | • Reliability (ICC) : 0.88-0.97<br>• Sensitivity : N/A<br>• Specificity : N/A |

Legend : ICC= intra Class Correlation Coefficient ; SEM= Standard error of Measurement ; MDC= Minimal Detectable Change ; M= men ; W= women ; sec = seconds ; N/A= not available

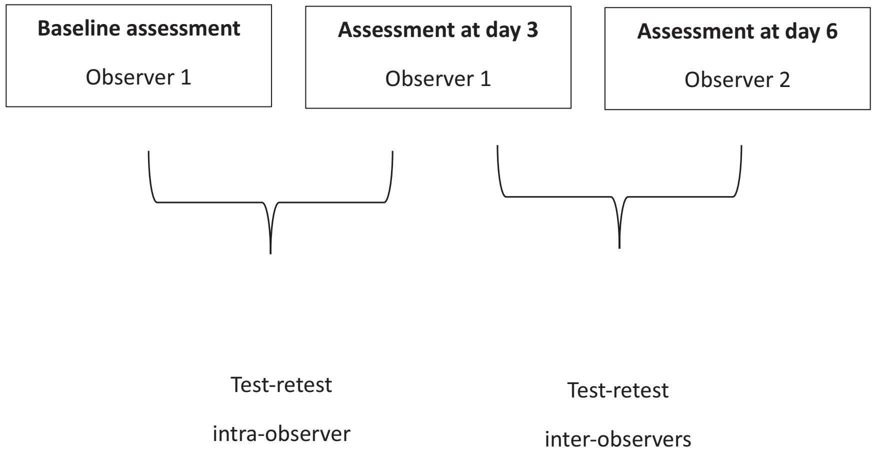


**Figure 1.** Organisation of data collection.

mean ± standard deviation (SD), or by median and interquartile range (P25-P75) for asymmetric distributions. Qualitative variables were expressed as numbers and percentages. The relative reliability of test−retest performed by the same observer or by two different ones was assessed using the intraclass correlation coefficient (ICC). The closer the coefficient is to 1, the higher is the reliability. We considered an ICC above 0.90 as very high, between 0.70 and 0.89 as high and between 0.50 and 0.69 as moderate.[25,26] Absolute reliability was examined using standard error of measurement [SEM = $\sqrt{}$(mean squared error)] and minimal detectable change (MDC = SEM*1.96*$\sqrt{2}$) [27]. SEM% and MDC% were calculated by dividing their respective values with the related average of the test and retest values. A relative SEM (%) with a value < 10% is considered acceptable in clinical practice. [27, 28] A small MDC(%) means a more sensitive measurement.[27,29] All calculations were performed using R statistical packages (Revolution Analytics, Redmond, WA, USA). Results were considered to be statistically significant at the 5% critical level ($P<0.05$).

## Results

### *Characteristics of the participants*

The mean age of the subjects was 77.7±7.7 years and 48.9% (n=22) of them were women. The mean BMI was 25.4±4.2 kg/m$^2$. The median socio-economic position index was 85 (73-115). 22 participants (48.9%) lived in France and 23 (51.1%) lived in Belgium. The mean scores obtained for the MMSE test and the Barthel Index were 30 and 100 respectively, which are the maximum scores that can be obtained for these two assessments (Table 2).

**Table 2**
Characteristics of the population (n=45)

| Characteristics | Mean ± SD Or Median (P25-P75) | N (%) |
|---|---|---|
| **Socio-demographic characteristics** | | |
| Age (years) | 77.7±7.7 | |
| Sexe (women:men) | | 22 (48.9):23 (51.1) |
| BMI (kg/m$^2$) | 25.4 ± 4.2 | |
| Socio-economic position index | 85 (73-115) | |
| Country of residence: | | |
| Belgium | | 23 (51.1) |
| France | | 22 (48.9) |
| MMSE score (/30) | 30 | |
| Barthel Index (/100) | 100 | |

The scores obtained in physical performances tests, performed in face-to-face and remotely by the 2 observers (day 3 and day 6, respectively) are presented in Table 3. Normal walking speed was 5.0 (4.3-6.2) second in face-to-face condition, and respectively 5.1 (4.4-5.9) and 5.3 (4.5-6.2) seconds in remote assessment (first and second observer).

### *Relative reliability of remote compared to face-to-face measurements*

#### *Intra-observer relative reliability*

The relative reliability of the remote assessments performed by the first observer compared with the assessments performed by the same observer in face-to-face mode was assessed through the ICC. The reliability was very high for our main criterion, the Normal Walking Speed (ICC = 0.96 (95% CI: 0.93-0.98)). The highest relative reliability values were observed for the Stand Test (ICC = 0.97 (95% CI: 0.95-0.99)) and the 10 repetitions Sit To Stand Test (ICC = 0.97 (95% CI: 0.94-0.98)) while the lowest value was observed for the 2-minute Step Test (ICC = 0.85 (95% CI: 0.69-0.92)) (Table 4).

#### *Inter-observer relative reliability*

The relative reliability of the distance assessments performed by the first observer compared with those performed by the second observer was assessed through the ICC. Relative reliability was considered very high for our main criterion, the Normal Walking Speed (ICC = 0.98 (95% CI: 0.96-0.99)).

The highest relative reliability value was observed for the 10 repetitions Sit To Stand Test (ICC = 0.99 (95% CI : 0.98-0.99)) while the lowest score was observed for the Fast walking speed (ICC =0.91 95% CI: 0.85-0.95) (Table 5).

### *Absolute reliability of remote measurements*

#### *Intra-observer absolute reliability*

Absolute intra-observer reliability was assessed through SEM% and MDC% values. The SEM% values ranged from 24.03% (Unipodal balance test) to 0% (Stand And Reach flexibility test). The Normal Walking speed (SEM%=12.30) was considered unacceptable from a clinical point of view (SEM%>10%). The MDC% values ranged from 9.93% (Unipodal balance test) to 0% (Stand And Reach flexibility test). Stand And Reach flexibility test is the most sensitive measure for detecting minimal change (Table 4).

**Table 3**
Results of Face-to-Face and Remote Physical Performance Assessments (n=45)

| | Face-to-Face Assessments (observer 1) | Remote Assessments (observer 1) | Remote Assessments (observer 2) |
|---|---|---|---|
| NWS (m/s) | 5.0 (4.3-6.2) | 5.1 (4.4-5.9) | 5.3 (4.5-6.2) |
| FWS (m/s) | 3.1 (2.6-3.8) | 3 (2.5-3.5) | 3.2 (2.8-3.8) |
| UBT (sec) | 8.4 (4.2-23.7) | 10.1 (3.7-27.9) | 13.9 (4.4-27.8) |
| NTUG (sec) | 10.3 (8.8-12.9) | 10.6 (8.7-12.4) | 10.4 (8.9-11.4) |
| FTUG (sec) | 6.7 (5.3-8.1) | 6.98 (5.1-8.6) | 6.8 (5.2-8.3) |
| 5STS (sec) | 10.9 (8.6-13.9) | 10.7 (8.6-15.1) | 10.4 (8.4-14.1) |
| 10STS (sec) | 25.5 (18.7-29.9) | 24.1 (18.3-31.4) | 23.9 (18.4-30) |
| 30sCS (number of repetitions) | 12 (11-17) | 12 (10-16.3) | 13.5 (10.8-16.3) |
| 2MST (number of repetitions) | 95.5 (84-126) | 88.5 (75.3-104.5) | 90 (75.8-108) |
| SAR (points) | 2 (1.75-3) | 2 (1.75-3) | 2 (1.75-3) |

Legend: NWS= Normal walking speed, FWS= Fast walking speed; UBT = unipodal balance test; NTUG = Normal Timed Up and Go; FTUG = Fast Timed Up and Go; 5STS = 5 repetitions sit to stand test; 10STS = 10 repetitions sit to stand test; 30CS = 30 seconds chair stand; 2MST = 2 minutes step test; SAD = Stand And Reach flexibility test

**Table 4**
Intra-observer relative and absolute reliability

| | ICC | 95% confidence interval | SEM | SEM (%) | MDC | MDC (%) |
|---|---|---|---|---|---|---|
| **NWS** | 0.96 | 0.93-0.98 | 0.75 | 12.30 | 0.31 | 5.0 |
| **FWS** | 0.88 | 0.80-0.93 | 0.79 | 21.04 | 0.32 | 8.6 |
| **UBT** | 0.93 | 0.88-0.96 | 3.62 | 24.03 | 1.49 | 9.93 |
| **NTUG** | 0.95 | 0.91-0.97 | 0.85 | 7.68 | 0.35 | 3.17 |
| **FTUG** | 0.91 | 0.85-0.95 | 1.00 | 13.44 | 0.41 | 5.56 |
| **5STS** | 0.97 | 0.95-0.99 | 0.79 | 6.58 | 0.33 | 2.72 |
| **10STS** | 0.97 | 0.94-0.98 | 1.47 | 5.72 | 0.61 | 2.37 |
| **30sCS** | 0.93 | 0.87-0.96 | 1.04 | 7.97 | 0.43 | 3.30 |
| **2MST** | 0.85 | 0.69-0.92 | 7.92 | 11.20 | 3.27 | 4.63 |
| **SAR** | 1 | 1-1 | 0 | 0 | 0 | 0 |

Legend: ICC= IntraClass Correlation coefficient; SEM (Standard Error of Measurement) = SD difference/√2; MDC (Minimal detectable Change) = SEM*1.96*√2; NWS= Normal walking speed, FWS= Fast walking speed; UBT = unipodal balance test ; NTUG = Normal Timed Up and Go; FTUG = Fast Timed Up and Go; 5STS = 5 repetitions sit to stand test; 10STS = 10 repetitions sit to stand test; 30CS = 30 seconds chair stand; 2MST = 2 minutes step test; SAR = Stand And Reach flexibility test.

**Table 5**
Inter-observer relative and absolute reliability

| | ICC | 95% confidence interval | SEM | SEM (%) | MDC | MDC (%) |
|---|---|---|---|---|---|---|
| **NWS** | 0.98 | 0.96-0.99 | 0.56 | 8.80 | 0.23 | 3.64 |
| **FWS** | 0.91 | 0.85-0.95 | 0.67 | 17.68 | 0.28 | 7.32 |
| **UBT** | 0.98 | 0.96-0.99 | 2.16 | 13.68 | 0.89 | 5.65 |
| **NTUG** | 0.92 | 0.86-0.96 | 1.18 | 10.60 | 0.49 | 4.38 |
| **FTUG** | 0.97 | 0.95-0.98 | 0.58 | 7.76 | 0.24 | 3.21 |
| **5STS** | 0.98 | 0.97-0.99 | 0.64 | 5.38 | 0.27 | 2.22 |
| **10STS** | 0.99 | 0.98-0.99 | 0.99 | 3.92 | 0.41 | 1.62 |
| **30sCS** | 0.95 | 0.90-0.97 | 0.84 | 6.44 | 0.35 | 2.66 |
| **2MST** | 0.92 | 0.86-0.95 | 6.66 | 7.61 | 2.75 | 3.14 |
| **SAR** | 1 | 1-1 | 0 | 0 | 0 | 0 |

Legend : ICC= IntraClass Correlation coefficient; SEM (Standard Error of Measurement) = SD difference/√2; MDC (Minimal detectable Change) = SEM*1.96*√2; NWS= Normal walking speed, FWS= Fast walking speed; UBT = unipodal balance test ; NTUG = Normal Timed Up and Go; FTUG = Fast Timed Up and Go; 5STS = 5 repetitions sit to stand test; 10STS = 10 repetitions sit to stand test; 30CS = 30 seconds chair stand; 2MST = 2 minutes step test; SAR = Stand And Reach flexibility test.

*Inter-observer absolute reliability*

Absolute inter-observer reliability was assessed using SEM% and MDC% values. The SEM% values ranged from 17.68% (Fast walking speed) to 0% (Stand And Reach flexibility test). The MDC% values ranged from 7.32% (fast walking speed) to 0% (Stand And Reach flexibility test). Stand And Reach flexibility test is the most sensitive measure for detecting minimal change (*Table 5*).

## Discussion

The main objective of this study was to investigate the reliability of remote physical performance assessment in older people when compared to traditional face-to-face evaluations. The results provided evidence to support the researchers' hypothesis. The absolute reliability of the remote assessments was rated as high or very high for all ten physical capacity tests, with ICCs ranging from 0.85 to 1 when compared to face-to-face assessments. However, the relative reliability was unacceptable for some tests, with SEM% and a MDC% of less than 10%. In particular, our main criterion, normal walking speed, showed a high relative reliability, with ICCs of 0.96 (0.93-0.98) and 0.98 (0.96-0.99) in the intra- and inter-observer conditions, respectively. The absolute reliability was also acceptable for normal walking speed in the inter-observer condition (SEM%=8.8 and MDC%=3.6), but unacceptable in the intra-observer condition (SEM%=12.3 and MDC%=5). These findings emphasize the importance of considering both relative and absolute reliability when evaluating the suitability of normal walking speed as a measure in different clinical or research contexts.

More in details, results showed that the (fast and normal) Timed Up and Go test (TUG), the 5 rep sit to stand test (5STS), and the unipedal balance test, all of which predictive of fall risk, [30] had a very high inter-observers/intra-observer relative reliability (ICC>0.90). In addition to the ICC analysis, the study used the SEM to assess the absolute reliability. Both the normal Timed Up and Go and the 5STS tests were considered acceptable in clinical routine for intra-observer reliability (SEM%<10%).[27,28] In inter-observer condition, the reliability was acceptable for the fast TUG and the 5STS (SEM<10%).[27,28] Finally, the MDC % values for all tests were below 10% (in intra- and inter- observer conditions), confirming the absolute reliability of the measurements.[29] These results suggest that the remote assessments are robust and valid, making them suitable for routine clinical assessment. When comparing our results with those of the study by Peyrusqué et al., [14] the intra-observer reliability values were very high for the fast TUG and the 5 STS (ICC>0.90), which corroborates our values. However, the intra-observer reliability value was high (ICC>0.70) for the normal TUG and the unipedal balance test in the Canadian study whereas the values in our study were very high. This difference can be partly explained by the contextual factors (who was conducting the assessments and how they were trained) and by the type of assessment used remotely (e.g. outdoor in the Canadian study because of the covid-19 restrictions [14] vs. at the homes of participants in our study). Anyway, the values were acceptable in clinical practice in both studies. Then, our results confirm those of Peyrusqué et al. regarding the SEM% values which were <10% for the normal Timed Up and Go and the 5STS tests. Note that the inter-observer reliability was not studied in the Canadian study. This measure was therefore an added value provided by our research.

For the 30 seconds chair stand test (30sCS), to assess muscle endurance, [31] the intra- and inter-observers relative reliability were very high in our study as well as in the study by Peyrusqué et al. (ICC>0.90).[14] In addition, the SEM % values were <10% in both studies. Our results confirm, on a larger scale and in Europe, the results obtained in Canada.

Regarding, the NWS, the FWS, the 2MST and the SAR test, which assess the physical abilities of older people, [31] the intra-and inter-observers relative reliability were high or very high. In inter-observer condition, only the SAR test was clinically acceptable (SEM%<10). However, in inter-observer condition, only the FWS test was not clinically acceptable (SEM%>10). The MDC% values were below 10%. Of these tests, only the NWS and FWS were used in the study by Peyrusqué et al.[14] The relative intra-observer reliability values obtained in this Canadian study were moderate and high. The absolute reliability was acceptable for the NWS (SEM%=9.3) but not for the FWS (SEM%=12.9) according to Peyrusqué et al.[14] The MDC% were above 25% for these tests.[14] In addition, Russell et al. obtained very high intra- and inter-observer values for the step test, [32] which is in line with our results.

The study's results aligned with previous research on remote assessments of physical capacity in older people regarding the reliability. Our main criteria, NWS, was considered valid, reliable and acceptable in clinical practice in Europe and North America. This remote test could therefore be a good alternative to in-person testing when the latter is not possible. In fact, the use of this remote NWS test would make it possible to identify the risk of loss of functional capacity and thus prevent loss of independence and reduced quality of life [33] or to implement appropriate public health measures to maintain functional capacity of older adults.[34] However, other tests have shown absolute reliability to be unacceptable for clinical use (MDC%>10) in our study (i.e. FWS and 2 MST in intra-observer condition). These tests must therefore be interpreted with caution since reliability had been demonstrated by other authors. A recent systematic review revealed that the correlation between face-to-face and remote measures of physical performance was inconsistent across studies.[35]

Surprisingly, our results consistently showed better performance in inter-observer measurements compared to intra-observer measurements across all the tests. Although the values were close in both conditions, several factors contribute to this intriguing difference. Firstly, it is noteworthy that as the same observer conducts repeated measurements in the intra-observer scenario, there exists the potential for familiarity to breed complacency. This familiarity may lead to a subtle reduction in the observer's attention to detail, resulting in decreased reproducibility over time. Furthermore, an additional explanation lies in the phenomenon of a "learning effect." When analyzing the inter-observer data, it is essential to recognize that this marks the third occasion on which the patient has undergone the same test. With repeated exposure, patients tend to grow more accustomed to the test's procedures and nuances. This growing familiarity enables them to refine their test-taking skills through practice, ultimately developing a routine that enhances their performance.[36]

When conducting assessments at participants' homes, several barriers and facilitators should be considered. Barriers may include issues related to space and configuration (e.g. start and stop line on the floor for NWS), as participants may face challenges if they lack sufficient space or their home environment is not conducive to creating suitable conditions for assessments. Additionally, distractions at home, such as background noise or interruptions, could compromise the quality of assessments.

On the other hand, several facilitators can help mitigate these barriers. Providing clear and detailed instructions to participants on how to prepare their environment and perform the assessments can simplify the process. Offering technical support to participants, such as online assistance, can help them troubleshoot any technology-related issues. Additionally, allowing flexibility in the tools used, such as letting participants choose between using a laptop or a mobile phone according to their preference, can enhance the user-friendliness of home assessments.

Caution should be exercised in interpreting the results of this survey as the study sample is not entirely representative of older individuals. In fact, the study only included individuals with internet access, while internet usage among older European individuals hovers at 45% (ranging from 14% in Greece to 80% in Sweden).[37] Furthermore, this study only focuses on older individuals with a considerable degree of independence, even though roughly 20% of seniors in Europe experience a decline in their autonomy.[38] Then, during the SAR test, the obtained score was perfect (ICC=1, SEM and MDC=0%). This could suggest a lack of sensitivity to change due to categorical scoring. A continuous scoring system would have been more appropriate to assess the validity and the reliability of a flexibility test. Regarding standardisation and measurement safety, remote evaluations encounter issues such as auditory problems that might affect the comprehension of instructions. Furthermore, digital technology often experiences "lags" due to weak internet connections, which can cause delays in the transmission of directions or counts remotely and subsequently lead to measurement inaccuracies. Such delays could substantially impact the assessment results, particularly for tests such as TUG, where even a 1.5-second variation could indicate a meaningful improvement or deterioration in performance.[39] A limiting factor to consider is the limited space in the patient's home, which may impede remote assessments despite having all necessary physical, cognitive, and technological elements in place. From a safety standpoint, some participants asked for a companion to be present during the balance test in remote assessments, suggesting that this test may not be entirely appropriate for all individuals. Having a partner present during the evaluations could introduce a competitive element that might compromise the standardization and neutrality established by the evaluator.

To enhance the study, it may be worthwhile to consider including a wider older population by using less restrictive selection criteria to be more representative of the older population. In fact, we only included only people without cognitive impairment (as assessed by the MMSE) and with a high level of autonomy (as assessed by the Barthel score). It may be useful to investigate various tests, such as the Berg Balance Test for balance and Senior Fitness Tests for flexibility and upper limb strength.

In conclusion, the study's results demonstrate high to very high intra and inter-observers relative reliability for the 10 remote tests used. In addition, results suggest that some of these remote evaluations could be considered in clinical practice (acceptable absolute reliability) for healthcare professionals assessing patients unable to visit in person. As healthcare continues to evolve in response to changing circumstances and technological advancements, the integration of remote assessments offers a promising avenue for improving patient care and expanding access to healthcare services. Further research and exploration of these remote evaluation methods are warranted to fully harness their potential and ensure their seamless integration into clinical practice.

## Funding

Fanny Buckinx is supported by the FNRS (Fonds National pour la Recherche Scientifique).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Fanny Buckinx:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Marvin Rezoulat:** Data curation, Formal analysis, Investigation, Writing – review & editing. **César Lefranc:** Data curation, Formal analysis, Investigation, Writing – review & editing. **Jean-Yves Reginster:** Investigation, Supervision, Writing – review & editing. **Olivier Bruyere:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

## References

1. Patrizio E, Calvani R, Marzetti E, Cesari M. Physical functional assessment in older adults. *J Frailty Aging*. 2021;10:141–149.
2. Haskell WL, Blair SN, Hill JO. Physical activity: health outcomes and importance for public health policy. *Prev Med*. 2009;49:280–282.
3. van Lummel RC, Walgaard S, Pijnappels M, et al. Physical performance and physical activity in older adults: associated but separate domains of physical function in old age. *PLoS One*. 2015;10: e0144048.
4. Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical perspective. *J Man Manip Ther*. 2011;19:91–99.
5. Pfeifer LO, Helal L, Oliveira NL, Umpierre D. Correlates of functional physical capacity in physically active older adults: a conceptual-framework-based cross-sectional analysis of social determinants of health and clinical parameters. *Aging Clin Exp Res*. 2023;35:127–136.
6. Usher R, Stapleton T. Assessing older adults' decision-making capacity for independent living: practice tensions and complexities. *J Appl Gerontol*. 2022;41:1264–1273.
7. Aubertin-Leheudre M, Rolland Y. The Importance of Physical Activity to Care for Frail Older Adults During the COVID-19 Pandemic. *J Am Med Dir Assoc*. 2020;21:973–976.
8. Granet J, Peyrusqué E, Ruiz F, et al. Online physical exercise intervention in older adults during lockdown: Can we improve the recipe? *Aging Clin Exp Res*. 2023:1–10.
9. Araujo de Carvalho I, Epping-Jordan J, Pot AM, et al. Organizing integrated health-care services to meet older people's needs. *Bull World Health Organ*. 2017;95:756–763.
10. Buckinx F, Waters DL, Aubertin-Leheudre M. Letter to the editor: implementing home-based exercise technology in a nursing home: does MCI status matter? *J Frailty Aging*. 2021;10:77–78.
11. Doraiswamy S, Jithesh A, Mamtani R, Abraham A, Cheema S. Telehealth use in geriatrics care during the COVID-19 Pandemic-A scoping review and evidence synthesis. *Int J Environ Res Public Health*. 2021;18.
12. Ding J, Yang Y, Wu X, Xiao B, Ma L, Xu Y. The telehealth program of occupational therapy among older people: an up-to-date scoping review. *Aging Clin Exp Res*. 2023;35:23–40.
13. Overend T, Anderson C, Sawant A, Perryman B, Locking-Cusolito H. Relative and absolute reliability of physical function measures in people with end-stage renal disease. *Physiother Can*. 2010;62:122–128.
14. Peyrusqué E, Granet J, Pageaux B, Buckinx F, Aubertin-Leheudre M. Assessing physical performance in older adults during isolation or lockdown periods: web-based video conferencing as a solution. *J Nutr Health Aging*. 2022;26:52–56.
15. Lauzé M, Martel DD, Aubertin-Leheudre M. Feasibility and effects of a physical activity program using gerontechnology in assisted living communities for older adults. *J Am Med Dir Assoc*. 2017;18:1069–1075.
16. Martel D, Lauzé M, Agnoux A, et al. Comparing the effects of a home-based exercise program using a gerontechnology to a community-based group exercise program on functional capacities in older adults after a minor injury. *Exp Gerontol*. 2018;108:41–47.
17. Granet J, Peyrusqué E, Ruiz F, et al. Online physical exercise intervention in older adults during lockdown: can we improve the recipe? *Aging Clin Exp Res*. 2023;35:551–560.
18. Granet J, Peyrusqué E, Ruiz F, et al. Web-based physical activity interventions are feasible and beneficial solutions to prevent physical and mental health declines in community-dwelling older adults during isolation periods. *J Gerontol A Biol Sci Med Sci*. 2023;78:535–544.
19. Buckinx F, Aubertin-Leheudre M, Daoust R, et al. Feasibility and acceptability of remote physical exercise programs to prevent mobility loss in pre-disabled older adults during isolation periods such as the COVID-19 Pandemic. *J Nutr Health Aging*. 2021;25:1106–1111.
20. Buckinx F, Aubertin-Leheudre M, Daoust R, et al. Impacts of remote physical exercises on functional status and mobility among community-dwelling Pre-disabled seniors during the Covid-19 lockdown. *J Nutr Health Aging*. 2023.
21. Shah S, Vanclay F, Cooper B. Improving the sensitivity of the Barthel Index for stroke rehabilitation. *J Clin Epidemiol*. 1989;42:703–709.
22. Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. *J Am Geriatr Soc*. 1992;40:922–935.
23. Beaudart C, Reginster JY, Petermans J, et al. Quality of life and physical components linked to sarcopenia: the SarcoPhAge study. *Exp Gerontol*. 2015;69:103–110.
24. Genoud PA. *Indice de position socioéconomique (IPSE): un calcul simplifié*. 2023. Université de Fribourg; 2011.
25. Bartko JJ. Measurement and reliability: statistical thinking considerations. *Schizophr Bull*. 1991;17:483–489.
26. Kolber MJ, Beekhuizen K, Cheng MS, Fiebert IM. The reliability of hand-held dynamometry in measuring isometric strength of the shoulder internal and external rotator musculature using a stabilization device. *Physiother Theory Pract*. 2007;23:119–124.
27. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19:231–240.
28. Cejudo A, Sainz de Baranda P, Ayala F, Santonja F. Test-retest reliability of seven common clinical tests for assessing lower extremity muscle flexibility in futsal and handball players. *Phys Ther Sport*. 2015;16:107–113.
29. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther*. 2006;86:735–743.
30. Shumway-Cook A, Baldwin M, Polissar NL, Gruber W. Predicting the probability for falls in community-dwelling older adults. *Phys Ther*. 1997;77:812–819.
31. Buckinx F, Peyrusqué E, Kergoat MJ, Aubertin-Leheudre M. Reference Standard for the Measurement of Loss of Autonomy and Functional capacities in long-term care facilities. *J Frail Aging*. 2023;12:236–243.
32. Russell TG, Hoffmann TC, Nelson M, Thompson L, Vincent A. Internet-based physical assessment of people with Parkinson disease is accurate and reliable: a pilot study. *J Rehabil Res Dev*. 2013;50:643–650.
33. Phelan EA, Ritchey K. Fall prevention in community-dwelling older adults. *Ann Intern Med*. 2018;169:Itc81–itc96.
34. Ethgen O, Beaudart C, Buckinx F, Bruyère O, Reginster JY. The future prevalence of sarcopenia in Europe: a claim for public health action. *Calcif Tissue Int*. 2017;100:229–234.
35. Heslop PA, Hurst C, Sayer AA, Witham MD. Remote collection of physical performance measures for older people: a systematic review. *Age Ageing*. 2023;52.
36. Wu G, Sanderson B, Bittner V. The 6-minute walk test: how important is the learning effect? *Am Heart J*. 2003;146:129–133.
37. Tavares AI. Self-assessed health among older people in Europe and internet use. *Int J Med Inform*. 2020;141: 104240.
38. Abdi S, Spann A, Borilovic J, de Witte L, Hawley M. Understanding the care and support needs of older people: a scoping review and categorisation using the WHO international classification of functioning, disability and health framework (ICF). *BMC geriatrics*. 2019;19:195.
39. Mesquita R, Wilke S, Smid DE, et al. Measurement properties of the Timed Up & Go test in patients with COPD. *Chron Respir Dis*. 2016;13:344–352.
40. Peters DM, Fritz SL, Krotish DE. 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*. 2013;36:24–30.

41. Gunasekaran V, Banerjee J, Dwivedi SN, Upadhyay AD, Chatterjee P, Dey AB. Normal gait speed, grip strength and thirty seconds chair stand test among older Indians. *Arch Gerontol Geriatr*. 2016;67:171–178.
42. Unver B, Baris RH, Yuksel E, Cekmece S, Kalkan S, Karatosun V. Reliability of 4-meter and 10-meter walk tests after lower extremity surgery. *Disabil Rehabil*. 2017;39:2572–2576.
43. Springer BA, Marin R, Cyhan T, Roberts H, Gill NW. Normative values for the unipedal stance test with eyes open and closed. *J Geriatr Phys Ther*. 2007;30:8–15.
44. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. *Gerontologist*. 2013;53:255–267.
45. Okabe T, Suzuki M, Goto H, et al. Sex differences in age-related physical changes among community-dwelling adults. *J Clin Med*. 2021;10.
46. Bohannon RW. Reference values for the timed up and go test: a descriptive meta-analysis. *J Geriatr Phys Ther*. 2006;29:64–68.
47. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc*. 1991;39:142–148.
48. Keller K, Engelhardt M. Strength and muscle mass loss with aging process. Age and strength loss. *Muscles Ligament. Tendons J*. 2013;3:346–350.
49. Bohannon RW. Reference values for the five-repetition sit-to-stand test: a descriptive meta-analysis of data from elders. *Percept Mot Skills*. 2006;103:215–222.
50. Fudickar S, Hellmers S, Lau S, Diekmann R, Bauer JM, Hein A. Measurement system for unsupervised standardized assessment of timed "Up & Go" and five times sit to stand test in the community-a validity study. *Sensors (Basel)*. 2020:20.
51. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol*. 1994;49:M85–M94.
52. Buatois S, Perret-Guillaume C, Gueguen R, et al. A simple clinical scale to stratify risk of recurrent falls in community-dwelling adults aged 65 years and older. *Phys Ther*. 2010;90:550–560.
53. Sawada S, Ozaki H, Natsume T, et al. The 30-s chair stand test can be a useful tool for screening sarcopenia in elderly Japanese participants. *BMC Musculoskelet Disord*. 2021;22:639.
54. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. *Res Q Exerc Sport*. 1999;70:113–119.
55. Nogueira MA, Almeida TDN, Andrade GS, et al. Reliability and accuracy of 2-minute step test in active and sedentary lean adults. *J Manipulat Physiol Ther*. 2021;44:120–127.
56. Srithawong A, Poncumhak P, Manoy P, et al. The optimal cutoff score of the 2-min step test and its association with physical fitness in type 2 diabetes mellitus. *J Exerc Rehabil*. 2022;18:214–221.
57. Langhammer B, Stanghelle JK. The senior fitness test. *J Physiother*. 2015;61:163.