Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Augmented reality-based walking test battery in young and older adults: intra-rater reliability and ecological validity

Flavie Le Rouzic
Université de Limoges
Justine Lacroix
Université de Limoges
Anthony Ehrmann
ILFOMER, Institut Limousin de Formation aux métiers de la réadaptation
Arnaud Boujut
Université de Limoges
Anaïck Perrochon
anaick.perrochon@unilim.fr

Université de Limoges

Research Article

Keywords: clinical walking test, complex walking, augmented reality, ecological validity, intra-rater reliability

Posted Date: June 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9912780/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of NeuroEngineering and Rehabilitation on October 8th, 2026. See the published version at https://doi.org/10.1186/s12984-026-02169-2.

Augmented reality-based walking test battery in young and older adults: intra-rater reliability and ecological validity

Flavie Le Rouzic[1], Justine Lacroix[1], Anthony Ehrmann[2], Arnaud Boujut[1,3], Anaïck Perrochon[1,2].

[1] Université de Limoges, HAVAE, UR 20217, F-87000 Limoges, France.

[2] ILFOMER, Institut Limousin de Formation aux métiers de la réadaptation, F-87000 Limoges, France.

[3] 3iL Ingénieurs, Limoges, France.

Contributing authors:

flavie.le-rouzic@etu.unilim.fr

justine.lacroix@unilim.fr

anthoehrmann@gmail.com

arnaud.boujut@unilim.fr

Corresponding author:

Anaïck Perrochon

Laboratoire HAVAE (UR20217)

123 Avenue Albert Thomas

87000 Limoges, France

anaick.perrochon@unilim.fr

Counts: Abstract: 247 words, Text: 2973 words, References: 45, Tables: 2, Figures: 3.

Abstract:

Background: Several walking tests assess mobility and fall-related function in older adults, but their clinimetric properties remain insufficiently documented across task conditions and assessment environments. Augmented reality (AR) may support standardized, instrumented assessment, but its ecological validity remains unclear.

Aim: This study assessed intra-rater reliability and ecological validity for a walking test battery in young and older adults.

Methods: Healthy young and older adults performed four clinical walking tests (straight walking, Figure-of-8 Walk, Timed Up and Go and obstacle crossing) in RW and AR, under single-task (ST) and arithmetic dual-task (DT) conditions. Completion time was used to assess within-session intra-rater reliability, agreement, and equivalence using predefined margins.

Results: Thirty-five young adults and 26 older adults were analysed. In young adults, AR-based test times showed good-to-excellent reliability and met RW equivalence in ST and DT conditions. In older adults, ST reliability was good to excellent for all tests, except obstacle crossing in AR (ICC = 0.719). Under DT conditions, reliability was moderate to good in both environments (ICC 0.698–0.862). In older adults, RW–AR equivalence under DT conditions was threshold-dependent: one DT test met the ±1.0 s criterion, whereas all met the ±2.0 s criterion.

Conclusions: AR-based walking assessment showed favourable clinimetric properties in young adults and under ST conditions in older adults. In older adults, DT performance showed greater variability and weaker RW–AR agreement. These findings support further development of standardized AR-based mobility assessment, while indicating that DT performance in older adults requires further study.

Keywords: clinical walking test, complex walking, augmented reality, ecological validity, intra-rater reliability.

Key points:

- AR-based walking times were reliable and agreed well with real-world performance in young adults.
- In older adults, AR–real-world agreement was stronger during single-task walking.
- Dual-task variability in older adults may reflect both reliability limits and adaptive strategies.
- Further validation is needed before AR walking tests can inform fall-risk assessment.

Introduction

Falls are a major public health issue and a key threat to functional independence in older adults [1]. Systematic fall-risk assessment may support the early identification of individuals at high risk and guide preventive interventions [1]. In clinical practice, fall risk is commonly assessed using standardized functional mobility tests, such as the Timed Up and Go (TUG) test or the Figure-of-8 Walk (F8W) [2–5]. Many of these tests involve complex walking, defined as non-linear walking that includes changes in direction and/or

rhythm [6], thereby placing greater demands on executive control [7, 8]. Dual-task (DT) versions of these tests [5, 9], further increase cognitive-motor interference and may reveal subtle deficits that are less apparent under single-task (ST) conditions, particularly in older adults [10]. Because such situations resemble the attentional and environmental demands of everyday mobility, they may offer greater ecological relevance for mobility assessment in ageing [11–13].

Despite their clinical relevance, the clinimetric properties of complex walking tests, particularly under DT conditions, remain incompletely documented [10]. The TUG is the most extensively studied test and shows excellent intra-rater reliability [14]. By contrast, evidence remains more limited for other walking paradigms. Under ST conditions, studies of the F8W in healthy older adults [5], and straight-walking (SW) in young and older adults have reported excellent reproducibility [15]. Under DT conditions, reliability has been examined in only a small number of studies, with generally good to excellent intra-rater reliability reported for DT versions of the F8W [5], SW and TUG [9]. A battery combining several walking tasks may provide a broader and more sensitive characterization of mobility, and may therefore improve the detection of subtle impairments relevant to fall risk [3, 4]. However, clinimetric data remain limited for many clinical walking tests, particularly under standardized ST and DT conditions. This gap limits the development and validation of multi-test batteries, despite their potential clinical value.

In routine care, the implementation of ST/DT walking test batteries is constrained by limited access to equipment, time pressure, protocol variability, and the difficulty of quantifying spatiotemporal gait parameters in real time [3, 16]. Extended reality

technologies, including augmented reality (AR) and virtual reality [17], may provide a practical framework for standardizing test administration while enabling automated capture of these parameters across clinical walking tests performed under ST and DT conditions [10]. Such an approach may be valuable in clinical assessment, as instrumented versions of the TUG have shown that moving beyond completion time alone to include spatiotemporal and kinematic parameters can improve the identification of fallers [18]. In parallel, studies using head-mounted display-based systems have reported good agreement for gait parameters measured during straight walking [19], the TUG [20, 21], and the 10-Meter Walk Test [22].

However, XR systems may influence locomotor behaviour and performance, making clinimetric evaluation essential before wider clinical use. Recent evidence indicates that user-experience factors, including comfort, sensory feedback and sense of presence, may shape cognitive and motor performance in immersive environments [23]. This issue is clinically relevant because immersive environments may alter locomotor behaviour compared with real-world (RW) walking [24]. Studies comparing identical walking tasks in RW and virtual reality environments have reported reduced velocity, increased step count, longer completion times and greater variability in virtual reality [25–27]. Yet evidence remains scarce for AR-based walking assessments [24]. To date, only one study in healthy young adults has examined within-session reliability of 8-m SW in AR across walking speeds, reporting intraclass correlation coefficient (ICC) values ranging from 0.861 to 0.93 [19]. In addition, only one AR study compared strategies used to negotiate real versus virtual obstacles over a short walkway (2.1 m), suggesting comparable behaviours [28]. To our knowledge, no study has examined an AR-based battery of clinically relevant walking tests under ST and DT conditions, nor its ecological

validity with equivalent RW performance in both young and older adults. Establishing such clinimetric evidence represents a necessary first step before AR-based walking assessments can be evaluated for clinical decision-making or fall-risk screening in older adults.

The aim of this study was to assess the intra-rater reliability of a battery of walking tasks performed under ST and DT conditions in RW and AR environments, and to assess ecological validity, in young and older adults. We hypothesized that temporal performance measures would show good intra-rater reliability overall, although greater variability might be observed under DT conditions because of the additional cognitive demands. We also hypothesized that AR performance would show agreement and equivalence with RW performance.

1. Materials and methods

1.1. Population

Two groups were enrolled in this comparative cross-sectional study: healthy young adults aged 18 to 30 years and healthy older adults aged 70 years or older. Young adults were recruited by email and word of mouth at the University of Limoges, whereas older adults were recruited from local associations and senior citizens' clubs in the Limoges area (Haute-Vienne, France). Participants were eligible if they were able to walk independently without an assistive device, understand instructions and provided written informed consent. Participants also had to report no neurological, orthopaedic, or cognitive disorders. Non-inclusion criteria were uncorrected visual impairment, implanted electronic medical devices (e.g., pacemaker), or use of hearing aids. Participants were excluded if they experienced a cybersickness symptoms during the

study, if their data were deemed unusable for technical reasons, or if they were unable to use the headset application. Ethical approval was obtained from the Ethics Committee for Research in Sport Sciences (Approval No. IRB00012476-2025-03-06-412) and was conducted in accordance with the Declaration of Helsinki.

1.2. Experimental procedure

Participants performed one SW test and three complex walking tests in both RW and in AR conditions (Figure 1.B.) i) 8m-SW, consisting of walking 8 m straight [1, 29] ; ii) the TUG consisting of standing up from a chair, walking 3m, turning around, walking back, and sitting down [30] ; iii) the F8W consisting of walking a figure-8 path around two cones spaced 1.5m [31] ; iv) 8 m obstacles crossing (OC), consisting of walking along an 8m walkway with four obstacles measuring 0.9 m x 0.14 m x 0.07 m, placed at intervals of 1.2 m, 0.8 m, 2.3 m, 1.7 m, and 2.0 m [32].

Each test was performed under both ST and DT conditions. The DT versions of the tests are hereafter referred to as SWDT, TUGDT, F8WDT and OCDT. In the DT condition, participants performed a serial subtraction by counting backwards in sevens [33]. The starting number was randomly generated between 101 and 201, excluding numbers ending in 7. For DT tests, participants were instructed to complete the walking task as quickly as possible while performing the cognitive task, without prioritizing either task.

Each test was repeated three times in each task condition and environment, yielding 24 trials per environment and 48 trials per participant (Figure 1.B.). Start and finish lines were predefined for all tests except the TUG, for which the chair defined the start and end position. Participants were asked to complete each test as quickly as possible.

Environment and test order were randomized using a Python script.

1.3. Experimental set-up

The experimental set-up is presented in Figure 1.A. AR assessments were performed using a Meta Quest® 3 mixed reality head-mounted display (Meta Platforms, Menlo Park, CA, USA) and a custom application using the Unity game engine (version 2022.3.31, Unity Technologies Inc., San Francisco, CA, USA). The application reproduced RW test conditions as closely as possible. For the TUG and F8W tests, the virtual cones were rendered opaque and were positioned at fixed spatial locations allowing participants to move freely around them (Figure 1.A.). These locations were defined relative to the participant's starting position, rather than being physically anchored. For the OC test, virtual obstacles were rendered transparent because opaque obstacles produced unnatural leg–object overlap during crossing. To standardize test initiation across trials, a starting zone was indicated by black footprints displayed on the floor. A positioning sound cue confirmed that the participant was correctly placed, whereas measurement start and end were marked by an audible signal when the participant crossed the start and finish lines. For the TUG, two calibrations were performed to detect sit-to-stand and stand-to-sit transitions: one at application launch to capture standing height and one during test setup to capture seated height.

Instructions were delivered through the application in written and audio formats, with experimenter assistance when necessary.

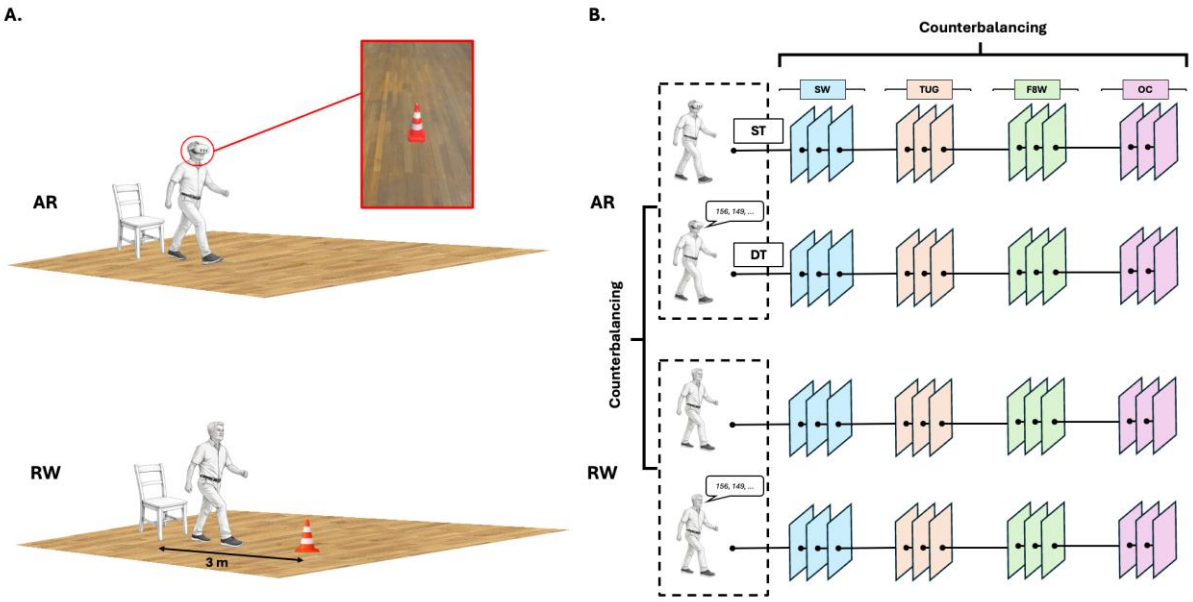


Figure 1. A. Experimental setup, **B.** Schematic diagram of the testing procedure for each participant.

Abbreviations: RW = real world; AR = augmented reality; ST = single-task; DT = dual-task; SW = straight-walk; F8W = figure-of-8 walk; OC = obstacle crossing; TUG = timed up and go; T = trial.

1.4. *Outcome*

The primary outcome was completion time. In the AR condition, time was recorded using the head-mounted display. In the RW condition, time was measured using a handheld stopwatch. Start and end times were determined using either the head-mounted display auditory cues described above or movement initiation, depending on the test.

1.5. Statistical analysis

Sample size was based on Geerse et al. (2020), which assessed a comparable AR device (HoloLens, Microsoft Corporation, Redmond, WA, USA) using Bland-Altman

analyses and ICC [19]. Their sample (23 young adults and 24 individuals with Parkinson's disease) reported high agreement (ICC ≥ 0.861). Using s = 0.2, an expected mean difference of 5% (d = 0.05), and 95% confidence, we estimated that 35 participants per group were required. Normality was assessed using the Shapiro-Wilk test (α = 0.05). Intra-rater reliability for completion time across the three trials was evaluated separately for each condition using the ICC for absolute agreement (ICC(A,1); [34]. ICC values above 0.50, 0.75, and 0.90 were interpreted as indicating moderate, good, and excellent agreement, respectively [35]. For ecological validity, the fastest time (best of three trials) was selected for each participant and condition. This choice was based on the common clinical use of best performance as an indicator of maximal functional capacity. Outliers were identified using the interquartile range (IQR) method (3.0 × IQR beyond Q1/Q3) [36]. Ecological validity was evaluated based on complementary approaches: (i) bias (AR – RW) and limits of agreement (LoA) derived from non-parametric Bland–Altman analysis [37]; and (ii) two one-sided paired-sample Wilcoxon signed-rank tests, with equivalence margins set at ±1.0 s and ±2.0 s for time. These margins were set at or below the smallest reported minimal detectable change for the TUG in community-dwelling older adults (2.08 s ; [38]. Holm's correction was applied to the p-values from the two one-sided paired-sample Wilcoxon signed-rank tests to account for multiple testing [39]. Analyses were performed in R (version 4.5.1) within RStudio (version 2025.09.0+387). Statistical significance was set at $p < 0.05$.

2. Results

2.1. Sample characteristics

A total of 69 participants were screened, including 35 young adults and 34 older adults. Eight older adults were excluded due to withdrawal (n = 3), vision impairment (n = 1), or unusable data (n = 4), leaving 26 participants for analysis. The final sample comprised 61 participants. The healthy young adult group had a median age of 22.0 years (Q1 = 21.0; Q3 = 23.5), with 18 women (51.4%). In the healthy older adult group, median age was 74.5 years (Q1 = 71.2; Q3 = 77.8), with 16 women (61.5%).

2.2. Intra-rater Reliability

Intra-rater reliability estimates for completion time are presented in Table 1. In young adults, ICC values indicated good to excellent reliability across all walking tests and environments, ranging from 0.817 to 0.935, except for the F8W in AR, which showed moderate reliability (ICC = 0.740). In older adults, reliability was more variable across tasks and environments. In the RW condition, all ST tests showed good to excellent reliability, with ICC values ranging from 0.878 to 0.955. In the AR condition, ST reliability was also generally good, although the OC test showed only moderate reliability (ICC = 0.719). Under DT conditions, reliability in older adults ranged from moderate to good in both environments, with ICC values between 0.698 and 0.862.

Table 1. Intra-rater reliability for completion time on tests in both environments for both groups.

	RW					AR				
Test		Trial 1	Trial 2	Trial 3			Trial 1	Trial 2	Trial 3	
	N	Med [Q1 – Q3]	Med [Q1 – Q3]	Med [Q1 – Q3]	ICC	N	Med [Q1 – Q3]	Med [Q1 – Q3]	Med [Q1 – Q3]	ICC

HYA	SW	35	3.83 [3.55 - 4.14]	3.84 [3.43 - 4.14]	3.85 [3.52 - 4.16]	0.923	35	4.22 [3.76 - 4.42]	4.06 [3.74 - 4.39]	4.02 [3.84 - 4.44]	0.918
	OC	35	4.14 [3.66 - 4.43]	4.2 [3.78 - 4.42]	4.14 [3.76 - 4.46]	0.935	35	4.72 [4.26 - 4.93]	4.47 [4.24 - 4.91]	4.41 [4.13 - 4.9]	0.852
	F8W	35	4.02 [3.64 - 4.3]	4.01 [3.52 - 4.4]	3.97 [3.5 - 4.32]	0.908	35	4.36 [3.92 - 4.93]	4.09 [3.62 - 4.46]	4.02 [3.49 - 4.28]	0.74
	TUG	35	4.72 [4.3 - 5.04]	4.69 [4.39 - 4.96]	4.59 [4.27 - 4.96]	0.897	34	4.32 [3.97 - 4.84]	4.26 [3.77 - 4.68]	4.11 [3.67 - 4.55]	0.891
	SWDT	35	4.82 [4.51 - 5.31]	4.92 [4.39 - 5.58]	5.06 [4.46 - 5.67]	0.825	33	5.5 [5.19 - 6]	5.49 [5.11 - 5.85]	5.5 [5.24 - 5.92]	0.92
	OCDT	35	5.12 [4.73 - 5.5]	5.29 [4.71 - 5.61]	5.34 [4.64 - 5.89]	0.817	35	6.02 [5.44 - 6.44]	5.75 [5.3 - 6.43]	5.93 [5.36 - 6.3]	0.911
	F8WDT	34	5.09 [4.47 - 5.5]	4.9 [4.51 - 5.73]	4.99 [4.32 - 5.47]	0.842	35	5.83 [5.18 - 6.72]	5.66 [5.09 - 6.39]	5.38 [5 - 6.22]	0.909
	TUGDT	35	5.69 [4.94 - 6.3]	5.54 [5.11 - 6.46]	5.69 [4.9 - 6.39]	0.876	35	5.71 [5.3 - 6.69]	5.78 [5.31 - 6.49]	5.97 [5.31 - 6.5]	0.893
HOA	SW	26	5.13 [4.47 - 5.46]	4.92 [4.62 - 5.32]	5.18 [4.75 - 5.43]	0.927	23	5.97 [5.33 - 6.56]	5.81 [4.86 - 6.76]	5.5 [4.79 - 5.89]	0.796
	OC	26	5.73 [5.1 - 6.3]	5.78 [5.09 - 6.3]	5.88 [4.86 - 6.57]	0.955	23	7.3 [6.04 – 8.18]	7.2 [6.23 – 7.49]	6.82 [6.14 – 7.47]	0.719
	F8W	26	6.27 [5.43 - 6.71]	5.77 [5.27 – 6.51]	5.59 [5.1 - 6.41]	0.878	22	6.39 [5.9 - 7.42]	5.94 [5.28 – 6.57]	5.58 [5.24 – 6.45]	0.751
	TUG	26	7.02 [6.2 - 7.93]	6.86 [6.07 - 7.66]	7.14 [6.35 - 7.77]	0.934	22	6.28 [5.7 - 7.35]	6.3 [5.32 - 7.42]	6.15 [5.38 - 7.03]	0.839
	SWDT	22	7.04 [6.75 – 8.04]	7.37 [6.68 - 8.1]	7.6 [6.74 - 9.15]	0.699	21	8.1 [7.44 – 11.24]	8.53 [6.82 – 10.1]	8.28 [7.32 – 10.03]	0.711
	OCDT	23	8.28 [7.29 - 9.34]	8.12 [7.45 - 9.21]	8.19 [7.01 – 9.77]	0.862	21	8.2 [7.78 – 10.08]	8.63 [7.89 - 10.25]	8.47 [7.75 – 9.76]	0.707
	F8WDT	25	8.24 [6.84 - 11.62]	9.22 [7.42 - 11.1]	8.04 [7.06 - 11]	0.743	24	9.23 [8.3 – 12.62]	8.97 [8.31 – 10.62]	9.05 [8.12 – 13.88]	0.855
	TUGDT	25	9.32 [7.92 – 10.31]	9.14 [8.16 - 9.82]	9.4 [8.38 - 10.42]	0.698	22	9.22 [8.43 – 11.81]	9.54 [8.43 - 10.82]	9.46 [8.24 – 11.19]	0.778

231

Abbreviations: RW = real world; AR = augmented reality; N = observation number; Med = median; Q1 – Q3 = first and third quartiles; ICC = intraclass correlation coefficient for absolute agreement; HYA = healthy young adults; HOA = healthy older adults, SW = straight-walk; F8W = figure-of-8 walk; TUG = timed up and go; OC = obstacles crossing; DT = dual-task.

2.3. Ecological validity

Agreement and equivalence analyses between RW and AR environments for completion time are presented in Table 2 and Figure 2. In both groups, median completion times were below 10 s for all tests. In young adults, median completion times in RW ranged from 3.69 s on SW to 5.50 s for TUGDT, whereas in AR they ranged from 3.96 s on F8W to 5.75 s for OCDT in AR. In older adults, median completion times, were consistently longer than in young adults, ranging from 4.72 s for SW to 8.77 s for TUGDT in RW, and from 5.39 s SW to 8.8 s for the TUGDT in AR. In young adults, all ST and DT tests met the predefined equivalence criteria between RW and AR, regardless of whether the ±1.0 s or ±2.0 s margin was applied. Under ST conditions, bias values ranged from -0.39 s for TUG to 0.39 s for OC, whereas under DT conditions, they ranged from 0.23 s for TUGDT to 0.79 s for OCDT. In older adults, the contrast between ST and DT conditions was more pronounced. All ST tests met the ±1.0 s equivalence criterion except OC ($p = 0.188$). Under ST conditions, bias values ranged from -0.33 s for TUG to 0.61 s for OC. By contrast, only one DT test met the ±1.0 s equivalence criterion (TUGDT), whereas all DT tests met the broader ±2.0 s criterion. Under DT conditions, bias values ranged from 0.08 s for TUGDT to 1.29 s for F8WDT. In both groups, limits of agreement were relatively consistent and narrower for ST tests. By contrast, DT tests showed wider and more variable limits of agreement across tasks, particularly in older adults (Figure 2).

Table 2. Ecological validity for completion time under single- and dual-task conditions in both groups.

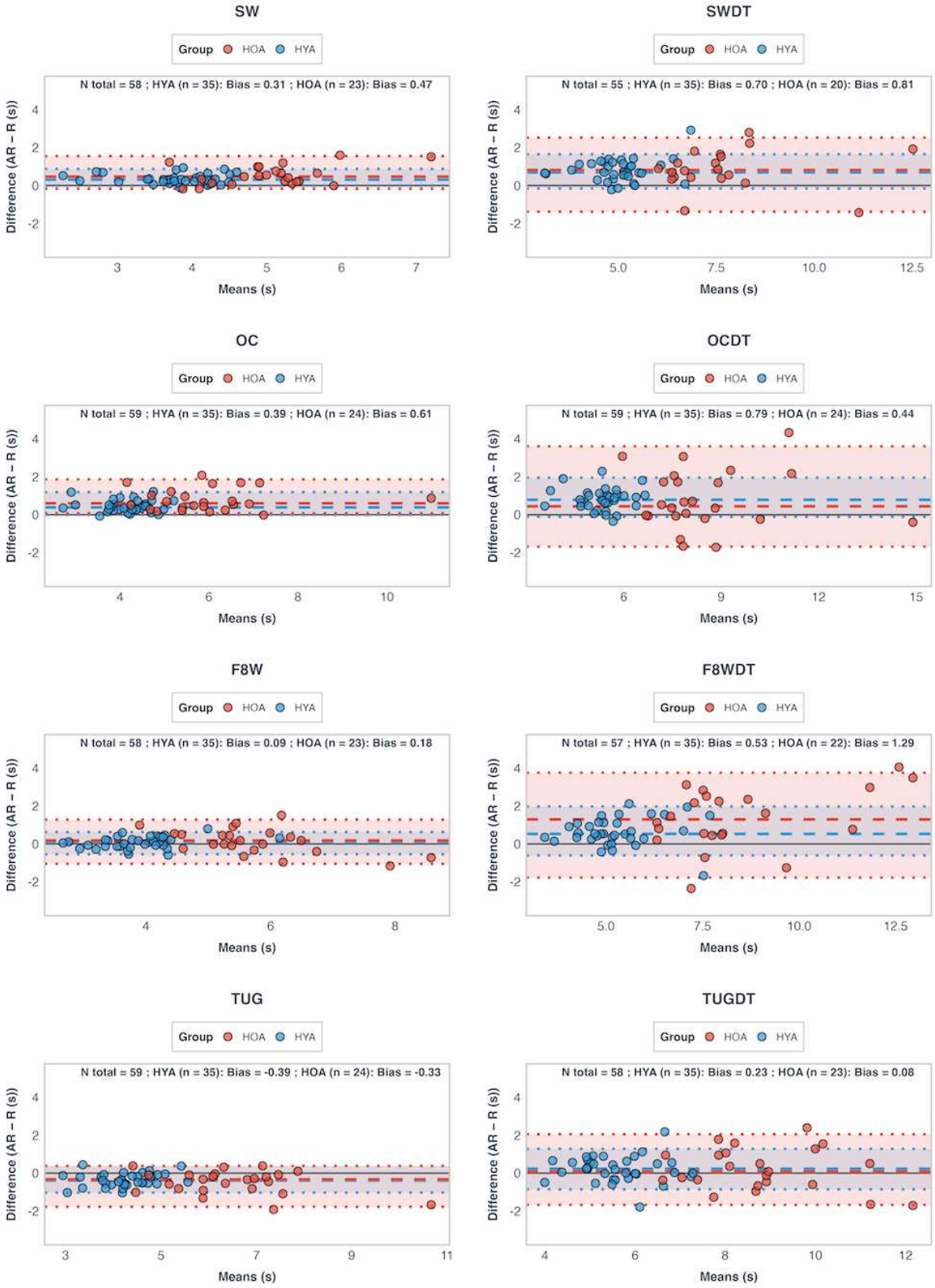
			RW	AR	Ecological validity statistics		
	Test	N	Med [Q1 – Q3]	Med [Q1 – Q3]	Bias [95% LoA]	Equivalence 1.0s margin	Equivalence 2.0s margin
HYA	SW	35	3.69 [3.43 - 4.02]	4.01 [3.73 - 4.35]	0.31 [0.00 ; 0.87]	0.004 **	0.004 **
	OC	35	4.03 [3.59 - 4.29]	4.41 [4.02 - 4.77]	0.39 [0.00 ; 1.19]	0.004 **	0.004 **
	F8W	35	3.89 [3.38 - 4.12]	3.96 [3.47 - 4.26]	0.09 [-0.55 ; 0.63]	0.004 **	0.004 **
	TUG	35	4.54 [4.16 - 4.86]	3.99 [3.64 - 4.54]	-0.39 [-1.03 ; 0.38]	0.004 **	0.004 **
	SWDT	35	4.7 [4.24 - 5.04]	5.44 [4.96 - 5.77]	0.70 [-0.15 ; 1.65]	0.02 *	0.004 **
	OCDT	35	5.06 [4.46 - 5.35]	5.75 [5.22 - 6.14]	0.79 [-0.11 ; 1.96]	0.02 *	0.004 **
	F8WDT	35	4.74 [4.31 - 5.36]	5.38 [4.98 - 6.05]	0.53 [-0.61 ; 1.98]	0.015 *	0.004 **
	TUGDT	35	5.5 [4.8 - 6]	5.52 [5.2 - 6.2]	0.23 [-0.85 ; 1.28]	0.004 **	0.004 **
HOA	SW	23	4.72 [4.39 - 5.23]	5.39 [4.74 - 5.53]	0.47 [-0.17 ; 1.56]	0.004 **	0.004 **
	OC	24	5.35 [4.73 - 6.18]	6.1 [5.48 - 6.88]	0.61 [0.08 ; 1.87]	**0.188**	0.004 **
	F8W	23	5.42 [4.97 - 6.02]	5.6 [5.24 - 6.39]	0.18 [-1.05 ; 1.28]	0.004 **	0.004 **
	TUG	24	6.62 [6.04 - 7.48]	6.11 [5.22 - 7.01]	-0.33 [-1.78 ; 0.38]	0.004**	0.004 **
	SWDT	20	6.92 [6.2 - 7.57]	8.02 [6.98 - 9.54]	0.81 [-1.38 ; 2.53]	**0.768**	0.024*
	OCDT	24	7.88 [6.8 - 8.66]	8.4 [7.57 - 9.75]	0.44 [-1.69 ; 3.61]	**0.768**	0.015*
	F8WDT	22	7.52 [6.21 - 9.34]	8.37 [7.61 - 9.45]	1.29 [-1.78 ; 3.76]	**0.851**	0.024*
	TUGDT	23	8.77 [7.46 - 9.37]	8.8 [8.26 - 10.46]	0.08 [-1.68 ; 2.05]	0.015*	0.004 **

Abbreviations: RW = real world condition; AR = augmented-reality condition; N = observation number; Med = median; Q1 – Q3 = first and third quartiles; LoA = limits of agreement; HYA = healthy young adults; HOA = healthy older adults, SW = straight-walk; F8W = figure-of-8 walk; TUG = timed up and go; OC = obstacles crossing; DT = dual-task.

*** Significant equivalence ($p < .001$); ** Significant equivalence ($p < .01$); * Significant equivalence ($p < .05$).

Values in bold represent non-significant values.

263


SW
Group HOA HYA
N total = 58 ; HYA (n = 35): Bias = 0.31 ; HOA (n = 23): Bias = 0.47
Difference (AR – R (s))
Means (s)
SWDT
Group HOA HYA
N total = 55 ; HYA (n = 35): Bias = 0.70 ; HOA (n = 20): Bias = 0.81
Difference (AR – R (s))
Means (s)
OC
Group HOA HYA
N total = 59 ; HYA (n = 35): Bias = 0.39 ; HOA (n = 24): Bias = 0.61
Difference (AR – R (s))
Means (s)
OCDT
Group HOA HYA
N total = 59 ; HYA (n = 35): Bias = 0.79 ; HOA (n = 24): Bias = 0.44
Difference (AR – R (s))
Means (s)
F8W
Group HOA HYA
N total = 58 ; HYA (n = 35): Bias = 0.09 ; HOA (n = 23): Bias = 0.18
Difference (AR – R (s))
Means (s)
F8WDT
Group HOA HYA
N total = 57 ; HYA (n = 35): Bias = 0.53 ; HOA (n = 22): Bias = 1.29
Difference (AR – R (s))
Means (s)
TUG
Group HOA HYA
N total = 59 ; HYA (n = 35): Bias = -0.39 ; HOA (n = 24): Bias = -0.33
Difference (AR – R (s))
Means (s)
TUGDT
Group HOA HYA
N total = 58 ; HYA (n = 35): Bias = 0.23 ; HOA (n = 23): Bias = 0.08
Difference (AR – R (s))
Means (s)

264

Figure 2. Bland-Altman plots of completion time agreement between environments under single- and dual-task conditions.

Abbreviations: RW = real world condition; AR = augmented-reality condition; N = observation number; HYA = healthy young adults; HOA = healthy older adults, SW = straight-walk; F8W = figure-of-8 walk; TUG = timed up and go; OC = obstacles crossing; DT = dual-task.

3. Discussion

By comparing AR-based walking tests with their RW counterparts under ST and DT conditions, this study evaluated intra-rater reliability and ecological validity for completion time in healthy young and older adults. Overall, reliability was high in young adults across tasks and environments. In older adults, reliability remained generally good under ST conditions but was more variable under DT conditions. Ecological validity was strongest in young adults and under ST conditions in older adults, whereas DT performance in older adults showed lower ecological validity and greater variability (Figure 3).

In RW, the selected walking tests showed high reproducibility in young adults and, under ST conditions, similarly strong reliability in older adults. These findings are consistent with the literature, as Bohannon (1997) reported excellent intra-rater reliability for SW in young adults [15]. For older adults in RW, ST results were similar to those observed in young adults, with high ICC values consistent with literature, as Exter et al. (2024) and Nualyong and Siriphorn (2022) documented excellent intra-rater reliability for the TUG and F8W [5,14]. In DT conditions, however, the increased task complexity was associated with greater variability compared with ST conditions, and, contrary to previous findings, reproducibility was only moderate [5, 9]. This discrepancy may reflect the concurrent cognitive task used, as serial subtraction, particularly by sevens, are among

289 the most demanding dual-task challenges and can markedly slow gait in older adults [40].
290 Participants had to shift their attention and redistribute limited attentional resources
291 between walking and calculation, which may have destabilized performance and reduced
292 reproducibility [41, 42]. Finally, this study provides initial reliability data for the obstacle-
293 crossing task under both ST and DT conditions in older adults.

294 In AR, reliability estimates broadly followed the same pattern as in RW, with generally
295 good reproducibility under ST conditions and greater variability under DT conditions in
296 older adults. Moreover, the present findings provide initial reliability data for an AR-based
297 walking test battery, including SW, TUG, F8W and OC. Overall, in older adults, DT
298 performance showed greater variability and moderate-to-good reproducibility across RW
299 and AR conditions.

300 Agreement analyses indicated that AR and RW completion times were generally
301 close under ST conditions, whereas agreement was weaker for DT assessments in older
302 adults**.** This favourable agreement under ST conditions contrasts with the virtual reality
303 literature, which frequently reports significant gait alterations [25–27], and reinforces the
304 advantage of AR in preserving the visibility of the physical environment [17]. By preserving
305 visual access to the physical environment, AR may help maintain natural sensorimotor
306 contingencies, which could partly explain the favourable agreement observed under ST
307 conditions [43, 44]. For older adults, the lack of equivalence likely reflects higher
308 between-trial variability induced by the concurrent cognitive task. This variability can be
309 viewed as a clinically meaningful expression of cognitive–motor interference [42, 45], and
310 of the strategies used by older adults to manage competing task demands [10]. However,
311 the present study does not allow the respective contributions of measurement instability,

cognitive–motor interference, and adaptive task-management strategies to be disentangled. The meaning of this variability therefore remains to be clarified.



		RW	AR	RW VS. AR
	ST	Excellent	Good	✓
	DT	Good	Excellent	✓
	ST	Excellent	Good	✓
	DT	Moderate	Moderate - Good	✗

Figure 3. Summary of intra-rater reliability and between-environment equivalence using the ±1.0s threshold across tasks.

Abbreviations: RW = real world condition; AR = augmented-reality condition; ST = single-task; DT = dual-task.

This study has several limitations. First, equivalence margins were derived from clinically meaningful thresholds reported for the TUG and were not task-specific, particularly under DT conditions. Second, the older-adult sample was smaller than planned, which may have reduced the precision of subgroup estimates. Third, older participants may have been influenced by unfamiliarity with the device, and cognitive-task performance was not analysed, limiting interpretation of cognitive–motor trade-offs.

Future studies should examine how age, dual-tasking, and AR-related task demands during complex walking, and should incorporate automated spatiotemporal measures and cognitive-task performance. Examining variability as a potentially informative outcome may also help determine whether DT-related fluctuations in AR reflect measurement noise, adaptive behaviour, or both, and whether they are associated with clinically relevant mobility profiles or fall-related outcomes. These findings provide foundational clinimetric evidence, but predictive validity for fall-risk screening remains to be established.

4. Conclusion

This study provides initial clinimetric evidence for an AR-based battery of clinical walking tests performed under single- and dual-task conditions. AR-derived completion time showed good intra-rater reliability and favourable ecological validity in young adults and in older adults under single-task conditions. In contrast, dual-task performance in older adults showed greater variability and weaker ecological validity, indicating that AR-derived dual-task measures should be interpreted cautiously at this stage. Overall, these findings support the further development of standardized, instrumented AR-based walking assessment in ageing and provide the clinimetric foundation required before its value for fall-risk screening and clinical decision-making can be tested.

List of abbreviations:

AR – Augmented Reality

DT – Dual-task

F8W – Figure-of-8 Walk

F8WDT – Dual-task Figure-of-8 Walk

ICC – Intraclass Correlation Coefficients

IQR – Interquartile Range

LoA – Limits of Agreement

OC – Obstacles Crossing

OCDT – Dual-task Obstacles Crossing

Q1 – First Quartile

Q3 – Third Quartile

RW – Real-world

ST – Single-task

SW – Straight-Walking

SWDT – Dual-task Straight-Walking

TUG – Timed Up and Go

TUGDT – Dual-task Timed Up and Go

Declarations:

Ethics approval and consent to participate: Ethical approval for this study was obtained from the Ethics Committee for Research in Sport Sciences, France (Approval No. IRB00012476-2025-03-06-412). In accordance with the Declaration of Helsinki, all subjects were informed about the purpose and procedure of the study and gave their informed consent to participate in the study.

Consent for publication: Not applicable.

Data availability: The datasets generated and analyzed during the current study are not publicly available due to ongoing analyses for additional manuscripts, but are available from the corresponding author on reasonable request.

Conflicts of interest: The authors declare no conflicts of interest.

Funding: This study received fundings from the Région Nouvelle-Aquitaine (Grant No. 32495610).

Author Contributions Statement: All authors contributed to the conception and design of the study. F.L. and A.E. performed the experiments and collected the data. F.L. carried out the data analysis and prepared the figures. F.L. and A.P. drafted the manuscript. J.L., A.E., A.B., and A.P. critically revised the manuscript. J.L. and A.P. supervised the research. All authors approved the final version of the manuscript.

Acknowledgements: We would like to thank all the volunteers who took part in this study, as well as the Master's intern Tristan DELAVAL for assisting with the recruitment of community-dwelling older adults and administering the assessments. We are also

grateful to 3iL Ingénieurs (School of Computer Engineering) for supporting the development of the application, providing access to its facilities, and enabling us to collaborate with its students as part of a project team. In particular, we would like to thank Axelle LÉGER for her contribution to the development of the augmented reality application.

References

1. Montero-Odasso M, Van Der Velde N, Martin FC, Petrovic M, Tan MP, Ryg J, et al. World guidelines for falls prevention and management for older adults: a global initiative. Age Ageing. Oxford University Press (OUP); 2022;51:1–36.

2. Beauchet O, Fantino B, Allali G, Muir SW, Montero-Odasso M, Annweiler C. Timed up and go test and risk of falls in older adults: A systematic review. J Nutr Health Aging. 2011;15:933–8.

3. Jepsen D, Robinson K, Ogliari G, Montero-Odasso M, Kamkar N, Ryg J, et al. Predicting falls in older adults: an umbrella review of instruments assessing gait, balance, and functional mobility. BMC Geriatr. 2022;22:615.

4. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta- analysis. BMC Geriatr. 2014;14:14.

5. Nualyong T, Siriphorn A. Accuracy of the figure of 8 walk test with and without dual-task to predict falls in older adults. J Bodyw Mov Ther. 2022;30:69–75.

404 6. Patla AE, Shumway-Cook A. Dimensions of Mobility: Defining the Complexity and
405 Difficulty Associated with Community Mobility. J Aging Phys Act. 1999;7:7–19.
406 https://doi.org/10.1123/japa.7.1.7

407 7. Hawkins KA, Fox EJ, Daly JJ, Rose DK, Christou EA, McGuirk TE, et al. Prefrontal over-
408 activation during walking in people with mobility deficits: Interpretation and functional
409 implications. Hum Mov Sci. 2018;59:46–55.

410 8. Mirelman A, Maidan I, Bernad-Elazari H, Shustack S, Giladi N, Hausdorff JM. Effects of
411 aging on prefrontal brain activation during challenging walking conditions. Brain Cogn.
412 2017;115:41–6.

413 9. Yang L, Liao LR, Lam FMH, He CQ, Pang MYC. Psychometric properties of dual-task
414 balance assessments for older adults: A systematic review. Maturitas. 2015;80:359–69.

415 10. Grosboillot N, Gallou-Guyot M, Lamontagne A, Bonnyaud C, Perrot A, Allali G, et al.
416 Towards a Comprehensive Framework for Complex Walking tasks: Characterization,
417 Behavioral Adaptations, and Clinical Implications in Ageing and Neurological
418 Populations. Ageing Res Rev. 2024;101:102458.

419 11. Dubbeldam R, Lee YY, Pennone J, Mochizuki L, Le Mouel C. Systematic review of
420 candidate prognostic factors for falling in older adults identified from motion analysis of
421 challenging walking tasks. Eur Rev Aging Phys Act Off J Eur Group Res Elder Phys Act.
422 2023;20:2.

12. Cai Y, Leveille SG, Andreeva O, Shi L, Chen P, You T. Characterizing Fall Circumstances in Community-Dwelling Older Adults: A Mixed Methods Approach. J Gerontol A Biol Sci Med Sci. 2023;78:1683–91.

13. Zukowski LA, Tennant JE, Iyigun G, Giuliani CA, Plummer P. Dual-tasking impacts gait, cognitive performance, and gaze behavior during walking in a real-world environment in older adult fallers and non-fallers. Exp Gerontol. 2021;150:111342.

14. Exter SH, Koenders N, Wees P, Berg MGA. A systematic review of the psychometric properties of physical performance tests for sarcopenia in community-dwelling older adults. Age Ageing. 2024;53:afae113.

15. Bohannon RW. Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. Age Ageing. 1997;26:15–9.

16. Hulleck AA, Menoth Mohan D, Abdallah N, El Rich M, Khalaf K. Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies. Front Med Technol. 2022;4:901331.

17. Rauschnabel PA, Felix R, Hinsch C, Shahab H, Alt F. What is XR? Towards a Framework for Augmented and Virtual Reality. Comput Hum Behav. 2022;133:107289.

18. Ortega-Bastidas P, Gómez B, Aqueveque P, Luarte-Martínez S, Cano-de-la-Cuerda R. Instrumented Timed Up and Go Test (iTUG)—More Than Assessing Time to Predict Falls: A Systematic Review. Sensors. 2023;23:3426.

19. Geerse DJ, Coolen B, Roerdink M. Quantifying Spatiotemporal Gait Parameters with HoloLens in Healthy Adults and People with Parkinson's Disease: Test-Retest Reliability, Concurrent Validity, and Face Validity. Sensors. 2020;20:3216.

20. Miller Koop M, Rosenfeldt AB, Owen K, Penko AL, Streicher MC, Albright A, et al. The Microsoft HoloLens 2 Provides Accurate Measures of Gait, Turning, and Functional Mobility in Healthy Adults. Sensors. 2022;22:2009.

21. van Bergem JS, van Doorn PF, Hoogendoorn EM, Geerse DJ, Roerdink M. Gait and Balance Assessments with Augmented Reality Glasses in People with Parkinson's Disease: Concurrent Validity and Test–Retest Reliability. Sensors. 2024;24:5485.

22. van Doorn PF, Geerse DJ, van Bergem JS, Hoogendoorn EM, Nyman E, Roerdink M. Gait Parameters Can Be Derived Reliably and Validly from Augmented Reality Glasses in People with Parkinson's Disease Performing 10-m Walk Tests at Comfortable and Fast Speeds. Sensors. 2025;25:1230.

23. Perrochon A, El Ayoubi K, Boujut A, Attoh-Mensah E. Beyond technology: How XR quality features shape user experience and physical activity determinants in exergaming. Acta Psychol (Amst). 2025;260:105764.

24. Sato K, Suda Y, Higuchi T. Assessing and enhancing the ecological validity of human locomotion in virtual reality. Front Virtual Real. 2026;6:1699143.

25. Mason AH, Padilla AS, Peer A, Toepfer M, Ponto K, Pickett KA. The role of the visual environment on characteristics of over-ground locomotion in natural and virtual environments. Int J Hum-Comput Stud. 2023;169:102929.

463 26. Nasiri M, Anaraky RG, Babu SV, Robb A. Gait Differences in the Real World and Virtual
464 Reality: The Effect of Prior Virtual Reality Experience. 2022 IEEE Int Symp Mix Augment
465 Real ISMAR. Singapore, Singapore: IEEE; 2022. p. 631–6.

466 27. Muhla F, Clanché F, Duclos K, Meyer P, Maïaux S, Colnat-Coulbois S, et al. Impact of
467 using immersive virtual reality over time and steps in the Timed Up and Go test in elderly
468 people. Anson ER, editor. PLOS ONE. 2020;15:e0229594.

469 28. Coolen B, Beek PJ, Geerse DJ, Roerdink M. Avoiding 3D Obstacles in Mixed Reality:
470 Does It Differ from Negotiating Real Obstacles? Sensors. 2020;20:1095.

471 29. Wade DT, Wood VA, Heller A, Maggs J, Langton Hewer R. Walking after stroke.
472 Measurement and recovery over the first 3 months. Scand J Rehabil Med. 1987;19:25–30.

473 30. Podsiadlo D, Richardson S. The Timed “Up & Go”: A Test of Basic Functional Mobility
474 for Frail Elderly Persons. J Am Geriatr Soc. 1991;39:142–8.

475 31. Hess RJ, Brach JS, Piva SR, VanSwearingen JM. Walking Skill Can Be Assessed in Older
476 Adults: Validity of the Figure-of-8 Walk Test. Phys Ther. 2010;90:89–99.

477 32. Chen H-C, Ashton-Miller JA, Alexander NB, Schultz AB. Stepping Over Obstacles: Gait
478 Patterns of Healthy Young and Old Adults. J Gerontol. 1991;46:M196–203.

479 33. Wollesen B, Wanstrath M, Van Schooten KS, Delbaere K. A taxonomy of cognitive tasks
480 to evaluate cognitive-motor interference on spatiotemoporal gait parameters in older
481 people: a systematic review and meta-analysis. Eur Rev Aging Phys Act. 2019;16:12.

482 34. McGraw KO, Wong SP. Forming inferences about some intraclass correlation
483 coefficients. Psychol Methods. 1996;1:30–46.

484 35. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation
485 Coefficients for Reliability Research. J Chiropr Med. 2016;15:155–63.

486 36. Tukey J. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley; 1977.

487 37. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat
488 Methods Med Res. 1999;8:135–60.

489 38. Donoghue OA, Savva GM, Börsch-Supan A, Kenny RA. Reliability, measurement error
490 and minimum detectable change in mobility measures: a cohort study of community-
491 dwelling adults aged 50 years and over in Ireland. BMJ Open. 2019;9:e030475.

492 39. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scand J Stat.
493 1979;6:65–70.

494 40. Hunter SW, Divine A, Frengopoulos C, Montero Odasso M. A framework for secondary
495 cognitive and motor tasks in dual-task gait testing in people with mild cognitive
496 impairment. BMC Geriatr. 2018;18:202.

497 41. Yogev-Seligmann G, Hausdorff JM, Giladi N. The role of executive function and
498 attention in gait. Mov Disord. 2008;23:329–42.

499 42. Al-Yahya E, Dawes H, Smith L, Dennis A, Howells K, Cockburn J. Cognitive motor
500 interference while walking: A systematic review and meta-analysis. Neurosci Biobehav
501 Rev. 2011;35:715–28.

502 43. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive
503 virtual environments. Philos Trans R Soc B Biol Sci. 2009;364:3549–57.

504 44. Sanchez-Vives MV, Slater M. From presence to consciousness through virtual reality.
505 Nat Rev Neurosci. 2005;6:332–9.

506 45. Plummer-D'Amato P, Brancato B, Dantowitz M, Birken S, Bonke C, Furey E. Effects of
507 Gait and Cognitive Task Difficulty on Cognitive-Motor Interference in Aging. J Aging Res.
508 2012;2012:1–8.

509