Karya J Health Sci. 2025;6(2):55-60
RESEARCH ARTICLE/ARAŞTIRMA MAKALESİ doi: 10.52831/kjhs.1640571 e-ISSN:2717-9540

KARYA JOURNAL OF HEALTH SCIENCE

journal homepage: www.dergipark.org.tr/kjhs

RELIABILITY OF TELE-ASSESSMENT OF PERFORMANCE-BASED TESTS IN PATIENTS WITH KNEE OSTEOARTHRITIS

DİZ OSTEOARTRİTLİ HASTALARDA PERFORMANS BAZLI TESTLERİN TELE-DEĞERLENDİRMESİNİN GÜVENİLİRLİĞİ

Gülşah Özsoy[1*], Hayriye Yılmaz[2], Nesibe Doğan[3]

[1]Department of Physiotherapy and Rehabilitation, Faculty of Health Sciences, Selçuk University, Konya, Türkiye

[2]Department of Physiotherapy and Rehabilitation, İzmir Bozyaka Education and Research Hospital, İzmir, Türkiye

[3]Clinic of Physical Medicine and Rehabilitation, İzmir Bozyaka Education and Research Hospital, İzmir, Türkiye

ABSTRACT

Objective: Knee osteoarthritis (OA) is one of the primary causes of disability, significantly impacting daily activities and quality of life. In recent years, tele-assessment methods have gained importance as effective tools for assessing physical performance, particularly for patients who face challenges in accessing in-person healthcare services. This study aimed to assess the intra- and inter-rater reliability of three commonly used performance-based tests - Timed Up and Go Test (TUG), Single Leg Stance Test (SLS), and Thirty-Second Chair Stand Test (30CST) - when applied in tele-assessment settings.

Method: A methodological study was conducted with 60 participants diagnosed with Grade 2-3 knee OA according to the Kellgren and Lawrence scale. Participants completed the TUG, SLS, and 30CST under two conditions: face-to-face in a clinical setting and remotely via tele-assessment (synchronized real-time video and asynchronized video recordings). Reliability was evaluated using Intraclass Correlation Coefficients (ICCs), Bland-Altman plots, and statistical measures of variability including Standard Error of Measurement (SEM) and Smallest Detectable Change (SDC).

Results: The inter-rater reliability between face-to-face and tele-assessment was good-to-excellent for TUG (ICC=0.824), SLS (ICC=0.902), and 30CST (ICC=0.848). Intra-rater reliability for tele-assessment was also good-to-excellent for TUG (ICC=0.949), SLS (ICC=0.814), and 30CST (ICC=0.926).

Conclusion: The findings indicate that TUG, SLS, and 30CST are reliable tools for tele-assessment in patients with knee OA. These tests can be effectively implemented in home-based telehealth programs to monitor physical performance and guide rehabilitation interventions. Their integration into remote care models may enhance patient engagement and adherence to rehabilitation protocols. Moreover, they provide clinicians with accessible and objective measures to make informed decisions regarding treatment progression.

Key Words: Knee Osteoarthritis, Functional Performance, Telehealth, Reproducibility of Results

ÖZ

Amaç: Diz osteoartriti (OA), engelliliğin başlıca nedenlerinden biri olup günlük aktiviteleri ve yaşam kalitesini önemli ölçüde etkilemektedir. Son yıllarda, özellikle yüz yüze sağlık hizmetlerine erişimde zorluk yaşayan hastalar için fiziksel performansın değerlendirilmesinde tele-değerlendirme yöntemleri giderek daha fazla önem kazanmıştır. Bu çalışma, tele-değerlendirme ortamında yaygın olarak kullanılan üç performans testinin -Zamanlı Kalk ve Yürü Testi (TUG), Tek Bacak Üzerinde Durma Testi (SLS) ve Otuz Saniyelik Sandalyeden Kalkma Testi (30SKT)- gözlemciler arası (inter-rater) ve gözlemci içi (intra-rater) güvenilirliğini değerlendirmeyi amaçladı.

Yöntem: Kellgren ve Lawrence skalasına göre Evre 2-3 diz OA tanısı konmuş 60 katılımcı ile yöntemsel bir çalışma gerçekleştirildi. Katılımcılar TUG, SLS ve 30SKT testlerini iki farklı koşulda tamamladılar: klinik ortamda yüz yüze ve senkron (gerçek zamanlı video); asenkron (kayıtlı video) tele-değerlendirme yoluyla uzaktan. Güvenilirlik değerlendirmesi için Sınıf İçi Korelasyon Katsayısı (ICC), Bland-Altman grafikleri ve Standart Ölçüm Hatası (SEM) ile En Küçük Algılanabilir Değişim (SDC) gibi değişkenlik istatistikleri kullanıldı.

Bulgular: Yüz yüze ve tele-değerlendirme arasındaki gözlemciler arası güvenilirlik, TUG (ICC=0.824), SLS (ICC=0.902) ve 30SKT (ICC=0.848) için iyi ile mükemmel arasında bulundu. Tele-değerlendirme içindeki gözlemci içi güvenilirlik ise TUG (ICC=0.949), SLS (ICC=0.814) ve 30SKT (ICC=0.926) için iyi ile mükemmel arasında değişti.

Sonuç: Bulgular, TUG, SLS ve 30SKT testlerinin diz OA olan hastalar için tele-değerlendirme ortamında güvenilir araçlar olduğunu göstermektedir. Bu testler, ev temelli tele-sağlık programlarında fiziksel performansı izlemek ve rehabilitasyon müdahalelerine rehberlik etmek için etkili bir şekilde uygulanabilir. Bu testlerin uzaktan bakım modellerine entegrasyonu, hastaların rehabilitasyon protokollerine katılımını ve uyumunu artırabilir. Ayrıca, klinisyenlere tedavi sürecinin ilerleyişi hakkında bilinçli kararlar verebilmeleri için erişilebilir ve objektif ölçümler sunar.

Anahtar Kelimeler: Diz Osteoartriti, Fonksiyonel Performans, Tele-Tıp, Sonuçların Tekrarlanabilirliği

Article Info/Makale Bilgisi

Submitted/Yükleme tarihi: 15.02.2025, **Revision requested/Revizyon isteği:** 30.04.2025, **Last revision received/Son düzenleme tarihi:** 17.05.2025, **Accepted/Kabul:** 22.05.2025

***Corresponding author/Sorumlu yazar:** Selçuk University, Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Konya, Türkiye
[1*]Email: fzt.gulsah@hotmail.com, [2]Email: fzthayriyeyilmaz@yahoo.com, [3]Email: nesibedogan@hotmail.com

INTRODUCTION

As the main cause of disability and loss of function in lower limb in middle-aged to older adults [1], knee osteoarthritis (OA) is the most common OA worldwide and its prevalence increases with age and obesity [2]. Among non-pharmacological methods, which are the first line of knee OA treatment [3], physical therapy is commonly prescribed to alleviate pain and improve physical function [4]. However, not all patients with knee OA have access to face-to-face rehabilitation sessions under direct supervision of physicians or physiotherapists [3].

Telemedicine has emerged as a viable alternative for delivering healthcare services remotely [5]. Tele-assessment, a key component of telemedicine, allows clinicians to evaluate patients' physical function using virtual platforms. Commonly used physical performance-based tests such as the Timed Up and Go Test (TUG), Single Leg Stance Test (SLS), and Thirty-Second Chair Stand Test (30CST) are widely accepted tools for assessing physical performance in patients with knee OA [6,7]. The TUG is a reliable test with acceptable minimum detectable change that can be clinically used in patients with doubtful to moderate knee OA [8]. The 30 CST is reliable in individuals with early-stage knee OA [9]. The Osteoarthritis Research Society International (OARSI) recommends the use of the TUG and the 30CST as reliable and valid tools for evaluating physical function in patients with hip and/or knee [10]. OA Likewise, the SLS is reliable in measuring centre of pressure during single-leg stance in individuals with knee OA [11]. In addition to these tests, the SLS is frequently used to assess balance, which are often compromised in individuals with lower limb OA. Impaired balance has been associated with an increased risk of falls and reduced mobility [12].

With the growing reliance on telehealth in healthcare delivery, it is essential to validate the psychometric properties of widely used performance-based tests in remote assessment settings. However, as telehealth becomes increasingly utilized in rehabilitation settings, it is essential to investigate whether these performance-based tests can be administered remotely with similar reliability. Factors such as lack of physical therapist supervision, limited home space, technology-related constraints, and patient safety concerns may influence test outcomes during tele-assessment. This study seeks to investigate the intra- and inter-rater reliability of the TUG, SLS, and 30CST when implemented through tele-assessment for patients with knee OA. By filling this critical gap in literature, our findings aim to enhance the evidence base supporting the integration of tele-assessment tools into remote-based rehabilitation programs.

METHOD

Study Design

This study was designed as a methodological study to evaluate the reliability of tele-assessment performance-based tests in patients with OA. Participants were assessed under two conditions: (1) in a clinical setting via face-to-face evaluations and (2) remote assessment using tele-assessment methods. Both synchronized (real-time) and asynchronized (video-recorded) tele-assessments were conducted.

Participants

We used convenience sampling method to recruit patients diagnosed with knee OA by their medical doctor according to the guidelines of American College of Rheumatology (ACR) [13]. Patient with (i) Grade 2 and 3 knee OA on the Kellgren and Lawrence (K/L) scale, radiologically confirmed [14], (ii) pain in the knee, and (iii) willingness to join the study were included.

Exclusion criteria included: (i) history of recent lower limb surgery, (ii) neurological or vestibular disorders affecting balance, and (iii) cognitive impairments interfering with test participation. We confirmed that all participants had access to internet connection and smartphones. Patients with (i) Grade 0, 1 and 4 knee OA on the K/L scale, (ii) central or peripheral nervous system involvement, (iii) previous knee surgery within the past six months, (iv) neurological or musculoskeletal disorders that would limit their performance on the tests, and (v) history of systemic arthritic conditions were excluded.

An a priori sample size calculation was conducted using the parameters for ICC analysis. Based on previous literature [15], we assumed an expected ICC of 0.75, a null hypothesis ICC of 0.50, a significance level (α) of 0.05, and a desired power of 0.80 with two raters. According to this analysis, a minimum of 60 participants was required to adequately detect reliability with these assumptions. Therefore, our study sample of 60 patients meets this requirement.

Outcome Measures

Sample Characteristics: Clinical and demographic data of the participants were recorded. Prior to testing, a specialist evaluated patients' radiological images to (1) assess lower extremity alignment and (2) assess disease severity using the K/L grading system. Frequently used in the literature, the K/L system checks the joint space narrowing and osteophytes [14]. In order to evaluate pain levels, functional capacity, and joint stiffness, we utilized the Turkish-adapted version of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [16], which is a valid, reliable, and responsive tool [17].

Timed up and go test (TUG): The TUG test is a straightforward and widely used functional assessment that evaluates an individual's mobility and balance. This test involves measuring the time required for a person to rise from a standard chair without using their arms for support, walk a distance of three meters at their usual walking speed, make a 180-degree turn, return to the starting position, and sit back down in the chair. Our participants performed 3 trials and their best score (in seconds) was recorded as the primary outcome [18]. Participants had at least 4 minutes of seated rest intervals between the trials.

Single Leg Stance Test (SLS): The participants used their leg with arthritic knee (in cases of unilateral involvement) or the more symptomatic leg (in cases of bilateral involvement) as the stance limb. The test was terminated when the free leg/foot touched the ground, or if excessive trunk/upper body movements (e.g., swinging arms) were observed. Participants were allowed to repeat the test if they scored less than10 seconds during the first attempt [19]. If needed, participants could rest between the trials for recovery.

Thirty second chair stand test (30 CST): The 30 CST checks the number of repetitions in 30 seconds that an individual can stand up from a standard chair (not using the arms/hands, feet flat on the ground) and sit back again. Our test chair was 43 cm in height and had no armrest. The participants performed 3 trials and their best score was recorded as the test result. If needed, they were allowed to rest (>5 min) between the trials for recovery [20].

Assessment Procedure

The participants were evaluated in two settings on different days (24-48 h apart): (i) in clinical settings (face-to-face) and (ii) remote via 'WhatsApp' phone application (both asynchronized and synchronized tele-assessment).

During the tele-assessment sessions, participants were given clear verbal instructions to perform all tests in a safe and controlled environment. For the SLS, participants were instructed to stand near a stable support surface, such as a wall, chair, or table, that they could hold onto if they lost balance. Before testing began, each participant was asked to confirm the presence of such safety support in their environment. Additionally, the evaluating therapist monitored the participant via video in real-time and was prepared to stop the test immediately if a safety concern was observed.

The evaluations steps were as follows:

1. Face-to-face/ Clinic assessment: Rater 1 completed all clinical assessments.
2. Synchronized tele-assessment: Rater 1 repeated the same evaluations via video calls to the patients (these were recorded to be used in step 3).
3. Asynchronized tele-assessment: Rater 2 evaluated the recordings of the remote evaluations (made in step 2).
4. Retest of synchronized tele-assessment: Rater 1 evaluated patients via video calls (these were recorded to be used in step 5).
5. Retest of asynchronized tele-assessment: Rater 2 evaluated the recordings of the remote evaluations (made in step 4).

The study follow-up is shown in Figure 1.

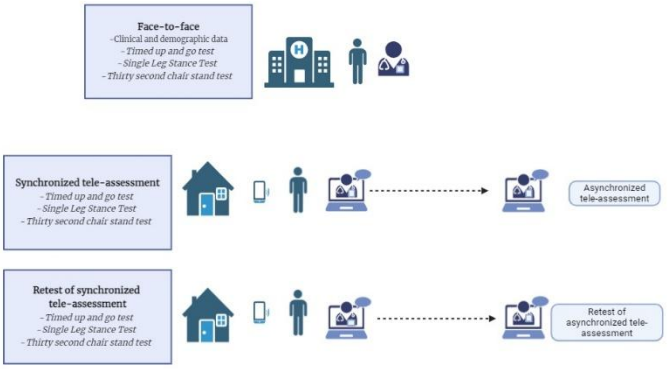

Figure 1. The scheme of the measuring procedure of the tests is as follows

Ethical Approval

All procedures performed in studies involving human participants strictly adhered to the ethical guidelines established by the appropriate institutional and/or national research committee. Furthermore, these procedures were conducted in full compliance with the ethical principles outlined in the Declaration of Helsinki, including its most recent revisions and any comparable internationally recognized ethical standards. Prior to the commencement of the study, ethical approval was obtained from the Ethics Committee of the University of Health Sciences İzmir Bozyaka Education and Research Hospital, ensuring that all aspects of the research met the highest ethical and regulatory requirements (date: 21.10.2020, approval number: 14). All participants gave written informed consent. The trial is registered under ClinicalTrials.gov with the clinical trial number NCT06711445.

Statistical Analysis

All statistical analyses in this study were conducted utilizing the most up-to-date version of the Statistical Package for Social Sciences (SPSS) version 25.0 software to ensure accuracy and reliability of the results. Descriptive statistical methods were employed to examine demographic and clinical data, with continuous variables expressed as mean ± standard deviation (SD) and categorical variables represented as percentages (%).

To assess potential systematic errors between different measurements, Bland-Altman plots were generated, providing a visual representation of the level of agreement between the two sets of values. Additionally, both intra-rater and inter-rater reliability were assessed through the calculation of Intraclass Correlation Coefficients (ICCs), a widely accepted statistical measure for evaluating consistency and reproducibility. The interpretation of ICC values was categorized as follows: scores belodobsw 0.75 were considered to indicate poor reliability, values ranging from 0.75 to 0.90 were classified as good reliability, and ICC scores exceeding 0.90 were deemed to reflect excellent reliability [21]. In addition to the reliability analyses, we also calculated the Standard Error of Measurement (SEM) to assess the extent of measurement variability. Furthermore, we determined the SEM with a 95% confidence interval (SEM95%) to provide a more precise estimate of the error associated with repeated measurements. Additionally, the Smallest Detectable Change at a 95% confidence level (SDC95%) was computed to identify the minimal amount of change that can be confidently distinguished from measurement error, ensuring the clinical relevance of the results [22].

RESULTS

Of the 67 recruited subjects, 7 were excluded (five patients had Grade 4 and two patients had Grade 1 knee OA). Therefore, the study was completed with a total of 60 participants (mean age: 56.43±7.31 years), of which 83.3% were females (n=50). Twenty-three participants had Grade 2, and 37 participants had Grade 3 knee OA. Table 1 presents the demographic and clinical data of the study sample. Participants completed all tests of the TUG, SLS, and 30 CST both in clinical settings and remote, indicating the absence of flooring effect for the tests. The test results and scores, including means and SDs, are shown in Table 2.

Table 1. Characteristics of the patients

Characteristic (n=60)	Mean (SD)
Gender, female, n (%)	50 (83.3)
Age, years	56.43±7.31
Height, cm	163.08±6.87
Weight, kg	81.83±14.01
Body mass index, kg/m^2	30.43±5.51
WOMAC score, %	39.28±4.23
K/L rating score, n (%)	
Grade 2	23 (38.3)
Grade 3	37 (61.7)
Limb involvement, n (%)	
Unilateral	35 (58.3)
Bilateral	25 (41.7)

Values are presented as mean ± standard deviation unless specified. WOMAC: Western Ontario and McMaster Universities Osteoarthritis index, K/L: Kellgren and Lawrence scale.

Reliability

The results demonstrated that the agreement between tele-assessment and face-to-face evaluation ranged from good to excellent for all functional tests examined. Specifically, inter-rater reliability was found to be strong for the TUG with an ICC value of 0.824, the SLS with an ICC of 0.902, and the 30CST with an ICC of 0.848. Additionally, when comparing the assessments performed by two different tele-raters, the inter-rater reliability was determined to be excellent across all tests, with ICC values of 0.963 for the TUG, 0.995 for the SLS, and 0.983 for the 30CST. A visual analysis of the Bland-Altman plots (Figures 2, 3, and 4) confirmed that no systematic trend indicating an improvement or decline in test performance was present, further supporting the consistency of the measurements. Similarly, the intra-rater reliability findings, including ICC95%, SEM, SEM95%, and SDC95%, are also provided in Table 3. The results indicated a good-to-excellent level of agreement for the synchronized tele-assessment method, with ICC values of 0.949 for the TUG, 0.814 for the SLS, and 0.926 for the 30CST. Consistent with these findings, the intra-rater reliability for the asynchronized tele-assessment method also demonstrated strong agreement, with ICC values of 0.925 for the TUG, 0.851 for the SLS, and 0.922 for the 30CST. Furthermore, Bland-Altman plot analysis (Figures 2,3, and 4) revealed no significant systematic variation in test performance, reinforcing the stability of the assessments over time.

Table 2. Range of data across assessments

Features	Clinic-Based Face-to-Face Assessment (Rater 1)	Remote -Based Synchronized Tele-Assessment (Rater 1)	Remote -Based Synchronized Tele-Assessment Retest (Rater 1)	Remote -Based Asynchronized Video Assessment (Rater 2)	Remote-Based Asynchronized video Assessment Retest (Rater 2)
TUG (s)	9.65 ± 2.23	9.93 ± 2.26	10.08 ± 2.19	9.86 ± 2.37	9.94 ± 2.22
SLS (s)	18.40 ±15.30	18.27±15.40	16.89±11.90	18.26±15.08	17.12±11.79
30CST(n)	9.25 ± 2.07	9.20 ± 1.86	9.31 ± 1.98	9.21 ± 1.86	9.35 ± 2.04

Data are presented as mean ± standard deviation. TUG: Timed Up and Go Test, SLS: Single Leg Stance Test, 30CST: Thirty Second Chair Stand Test

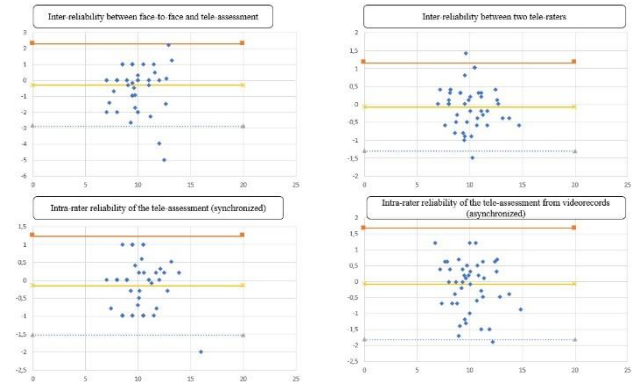

Figure 2. Bland-Altman plots for the Timed and Go Test

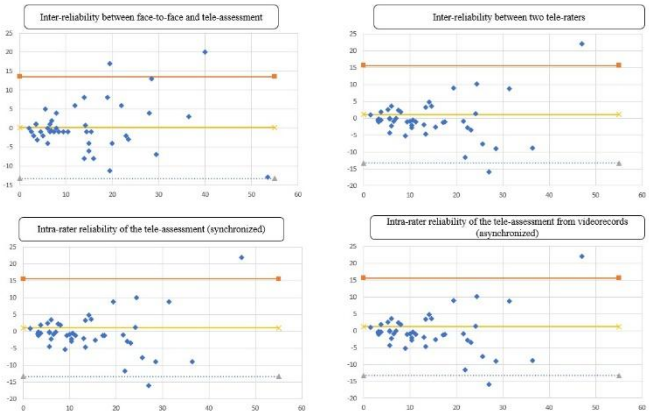

Figure 3. Bland-Altman plots for the Single Leg Stance Test

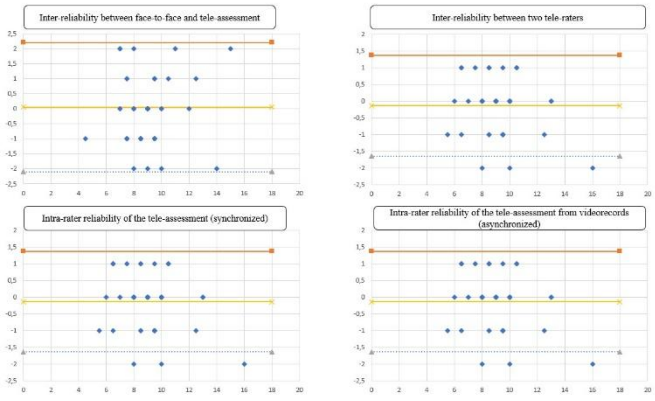

Figure 4. Bland-Altman plots for the Thirty Second Chair Stand Test

Additionally, statistical analyses using repeated measures of ANOVA indicated that there were no significant differences between the two synchronized tele-assessments for any of the functional tests.

The results showed that the TUG (F=2.593, p=0.113), the SLS (F=1.643, p=0.205), and the 30CST (F=1.497, p=0.226) exhibited consistent performance across repeated measurements. These findings further support the reliability of both tele-assessment methods, demonstrating their potential as accurate and reproducible alternatives to traditional face-to-face evaluations.

DISCUSSION

In this study, psychometric properties of the TUG, SLS, and 30 CST used as tele-assessment tests are investigated for the first time in patients with knee OA. With good-to-excellent intra- and inter-rater reliability, the TUG, SLS, and 30 CST tests can be reliably used as tele-assessment measures in patients with knee OA.

The demand for innovative telehealth services to reduce healthcare costs has grown over the past several years as a result of the globalization of healthcare systems. While the high cost of the technical equipment hindered the availability of telehealth in the 1990s [23], subsequent advancements in communication and technology have offered a large range of lower costs solutions for eHealth [24]. To perform tele-assessments in the current study, videocalls between the participants and the evaluator were made using WhatsApp software, which is a cheap and user-friendly program that can be used for health industry communications [25]. Recent studies in various clinical populations have demonstrated that tele-assessment methods yield reliable results in functional performance tests. High intra- and inter-rater reliability has been reported for tests such as the TUG and 5xSTS in patients with COPD [26], older adults [27], and individuals with chronic low back pain [28]. Similarly, balance assessments like the Berg Balance Scale and Dynamic Gait Index have shown strong agreement between face-to-face and remote evaluations [29]. These findings support the applicability of tele-assessment in different settings and reinforce the clinical feasibility of using tests like TUG, SLS, and 30CST for remote evaluation in patients with knee osteoarthritis.

The TUG test is recommended by the OARSI as a test of physical function in patients with hip or knee OA [10]. The TUG test has demonstrated good reliability and minimal detectable change in individuals with Grade 1–3 knee OA, supporting its use as a functional mobility assessment in this population [8]. To investigate measurement characteristics of performance-based tests of physical function in patients with hip and knee OA, Dobson et al. [30] conducted a systematic review. The authors reported that the TUG and 30CST were among the tests with the best measurement evidence for hip/knee OA [30]. To the best our knowledge, the psychometrics of the TUG as a tele-assessment test in patients with knee OA have not been studied before. Good inter-reliability (between face-to-face and tele-TUG) and excellent intra-reliability (between synchronized and asynchronized TUG) was confirmed in our study. The TUG is a convenient test for home settings as it needs no large space or equipment. The test is a common clinical balance test for people with knee OA [7]. The SLS test has been validated as a reliable measure of static balance and postural control in patients with knee OA [12].

Table 3. Reliability results of the tests

	Inter-reliability between face-to-face and tele-assessment				Inter-reliability between two tele-raters			
	ICC (95%CI)	**SEM**	**SEM95%**	**SDC95%**	**ICC (95%CI)**	**SEM**	**SEM95%**	**SDC95%**
TUG (s)	0.824 (0.722-0.891)	0.55	1.08	1.53	0.963 (0.940-0.978)	0.12	0.24	0.33
SLS (s)	0.902 (0.841-0.940)	2.14	4.20	5.94	0.995 (0.992-0.997)	0.10	0.20	0.29
30CST (n)	0.848 (0.758-0.906)	0.43	0.84	1.18	0.983(0.972-0.990)	0.04	0.08	0.12
	Intra-rater reliability of the tele-assessment (synchronized)				**Intra-rater reliability of the tele-assessment from video records (asynchronized)**			
	ICC (95%CI)	**SEM**	**SEM95%**	**SDC95%**	**ICC (95%CI)**	**SEM**	**SEM95%**	**SDC95%**
TUG (s)	0.949 (0.915-0.969)	0.16	0.31	0.44	0.925 (0.878-0.955)	0.24	0.48	0.68
SLS (s)	0.814 (0.708-0.884)	3.60	7.07	9.99	0.851 (0.763-0.908)	2.85	5.58	7.89
30CST (n)	0.926 (0.879-0.955)	0.20	0.39	0.56	0.922 (0.873-0.952)	0.21	0.42	0.60

ICC: Intraclass correlation coefficient, CI: Confidence interval, SEM: Standard error measurement, SDC: Smallest detectable change, TUG: Timed Up and Go Test, SLS: Single Leg Stance Test, 30CST: Thirty Second Chair Stand Test

We examined the psychometrics of the SLS as a tele-assessment test in patients with knee OA. Our results showed that the test has a high inter-rater reliability (between face-to-face and tele-SLS) and good intra-rater reliability (between synchronized and asynchronized SLS).

The 30CST is among the performance-based tests recommended by OARSI for individuals with knee OA [10]. The 30CST has also shown strong test-retest reliability and agreement in individuals with early-stage knee OA, making it a suitable tool for evaluating lower extremity strength and endurance [9]. Our results showed that the test can be reliably used in the same population as a tele-assessment method with high intra-rater and good inter-rater reliability.

The accuracy of repeated test scores is highly determined by the absolute reliability of the measurements and tests. Because the SEM and SDC95% are shown in the identical units as the instruments of measurement, they are more clinically useful than the ICC values. SEM is an indicator of the reliability of a measurement tool and expresses the amount of measurement error it contains. In clinical practices, the SDC95% can be used to categorize research sample participants as 'changed' or 'unchanged'. For instance, SEM95% for synchronized tele-TUG was 0.31s. This means that if a patient's TUG score is 10s, the patient's true score would -95% of the time- fall within the range of 9.69-10.31s. For the same test settings, SDC95% was 0.44s. This means that for a patient with a TUG score of 10s, any score between 9.56-10.44s on the repeated test would -95% of the time- represent a truly 'unchanged' performance. It should be noted that these values provide information about the measurement error and do not represent the minimal clinically important difference values [31].

In addition to findings from individuals with knee osteoarthritis, studies involving patients who have undergone total knee arthroplasty (TKA) also support the use of functional performance tests in both in-person and remote assessment settings. Yüksel et al. [32] reported excellent test-retest reliability for the TUG (ICC = 0.98), along with minimal detectable change (MDC) values that allow clinicians to interpret meaningful improvements in function following TKA. Similarly, Ünver et al. [33] demonstrated high reliability for the 30CST (ICC = 0.92) and the 50-Foot Walk Test (ICC = 0.97), confirming their utility in clinical evaluation after surgery. Additionally, Saraç et al. [34] validated the TUG, SLS, 2 Minute Walk Test (2MWT), and Five Times Sit-to-Stand Test (5xSST) as reliable and valid outcome measures for assessing balance in TKA patients. These results indicate that tele-assessment methods, including those used in the current study, may be reliably applied across the continuum of care from pre-operative OA management to post-operative rehabilitation.

The TUG, SLS, and 30CST test results contained no statistically significant systematic error, according to the repeated-measures ANOVA findings. This proved that the trials had no systematic difference.

This study demonstrates that the TUG, SLS, and 30CST can be reliably administered via tele-assessment in individuals with knee osteoarthritis. These findings support the use of remote functional testing as a practical option in cases where in-person assessments are not feasible, such as during pandemics or for patients with mobility or access limitations. The ability to perform these tests safely at home using minimal equipment offers significant potential for home-based rehabilitation and follow-up.

Future research should focus on the responsiveness of these tests to clinical change in tele-assessment settings and include broader patient populations with varying OA severity. Additionally, evaluating patient satisfaction, safety, and digital literacy may help improve the feasibility and effectiveness of remote assessments.

Limitations

This study has several limitations that should be acknowledged. Second, the generalizability of our findings is limited to patients with Grade 2 and 3 knee osteoarthritis. Individuals with more advanced (Grade 4) or very early (Grade 0–1) disease were excluded, and therefore, the results may not fully represent the entire spectrum of knee OA.

Second, the tele-assessments were conducted in participants' home environments, which inherently vary in terms of space, lighting, flooring, and noise levels. These environmental factors could potentially influence test performance and introduce variability compared to standardized clinical settings.

Third, in the asynchronized video-recorded assessments, although raters followed a standardized scoring protocol, the absence of real-time interaction may have limited their ability to clarify patient performance or correct errors during the tasks. This could lead to minor observer bias or reduced scoring accuracy.

Fourth, although body mass index (BMI) was recorded, its potential relationship with disease severity or functional performance was not examined in this study. Future research should investigate the impact of BMI on test outcomes in tele-assessment settings, as it may influence balance, mobility, and test performance.

Finally, we did not formally assess participants' technology literacy or comfort with using smartphones or video applications. Variations in digital familiarity may have affected test execution or communication during remote assessments, particularly among older adults.

CONCLUSION

The clinical implications of these findings are considerable, particularly in the context of increasing demand for remote healthcare solutions. The ability to reliably assess functional performance in patients with knee osteoarthritis through telehealth platforms offers clinicians a valuable means of maintaining continuity of care while minimizing the need for in-person visits. This is especially relevant for individuals with mobility limitations or those residing in remote areas. By integrating TUG, SLS, and 30CST into standard tele-rehabilitation protocols, healthcare providers can ensure consistent monitoring, early identification of functional decline, and timely adjustments to treatment plans. This not only enhances the quality of care but also promotes patient autonomy and engagement in their rehabilitation process.

Ethical Approval: *2020/14 Health Sciences University İzmir Bozyaka Education and Research Hospital Clinical Research Ethics Committee*

Conflict of Interest: *The authors have no conflicts of interest to declare.*

Funding: *None.*

Acknowledgements: *The authors sincerely appreciate the participation of all attendees and extend their gratitude for the time, effort, and valuable contributions provided.*

Author Contribution: Concept: *GO, HY, ND;* ***Design:*** *GO, HY, ND;* ***Data collecting:*** *GO, HY;* ***Statistical analysis:*** *GO;* ***Literature review:*** *GO, HY, ND;* ***Writing:*** *GO, HY;* ***Critical review:*** *GO, HY, ND.*

REFERENCES

1. Courties A, Kouki I, Soliman N, Mathieu S, Sellam J. Osteoarthritis year in review 2024: epidemiology and therapy. Osteoarthritis Cartilage. 2024;32(11):1397-1404.
2. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2024;403(10440):2133-2161.
3. Dantas LO, Salvini TF, McAlindon TE. Knee osteoarthritis: key treatments and implications for physical therapy. Braz J Phys Ther. 2021;25(2):135-146.
4. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage. 2019;27(11):1578-1589.
5. Xie SH, Wang Q, Wang LQ, Wang L, Song KP, He CQ. Effect of internet-based rehabilitation programs on improvement of pain and physical function in patients with knee osteoarthritis: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res. 2021;23(1):e21542.
6. Chen H, Wang C, Wu J, et al. Measurement properties of performance-based measures to assess physical function in knee osteoarthritis: A systematic review. Clin Rehabil. 2022:2692155221107731.
7. Hatfield GL, Morrison A, Wenman M, Hammond CA, Hunt MA. Clinical tests of standing balance in the knee osteoarthritis population: systematic review and meta-analysis. Phys Ther. 2016;96(3):324-337.
8. Alghadir A, Anwer S, Brismée JM. The reliability and minimal detectable change of timed up and go test in individuals with grade 1-3 knee osteoarthritis. BMC Musculoskelet Disord. 2015;16:174.
9. Gill S, Hely R, Page RS, Hely A, Harrison B, Landers S. Thirty second chair stand test: test-retest reliability, agreement and minimum detectable change in people with early-stage knee osteoarthritis. Physiother Res Int. 2022;27(3):e1957.
10. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage. 2013;21(8):1042-1052.
11. Takacs J, Carpenter MG, Garland SJ, Hunt MA. Test re-test reliability of centre of pressure measures during standing balance in individuals with knee osteoarthritis. Gait Posture. 2014;40(1):270-273.
12. Hunt MA, McManus FJ, Hinman RS, Bennell KL. Predictors of single-leg standing balance in individuals with medial knee osteoarthritis. Arthritis Care Res (Hoboken). 2010;62(4):496-500.
13. Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum. 1986;29(8):1039-1049.
14. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957;16(4):494-502.
15. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998;17(1):101-110.
16. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15(12):1833-1840.
17. Tüzün EH, Eker L, Aytar A, Daşkapan A, Bayramoğlu M. Acceptability, reliability, validity and responsiveness of the Turkish version of WOMAC osteoarthritis index. Osteoarthritis Cartilage. 2005;13(1):28-33.
18. Podsiadlo D, Richardson S. The timed "up & go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142-148.
19. Takacs J, Garland SJ, Carpenter MG, Hunt MA. Validity and reliability of the community balance and mobility scale in individuals with knee osteoarthritis. Phys Ther. 2014;94(6):866-874.
20. Gill S, McBurney H. Reliability of performance-based measures in people awaiting joint replacement surgery of the hip or knee. Physiother Res Int. 2008;13(3):141-152.
21. Koo TK, Li MY. A Guideline of selecting and reporting ıntraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-163.
22. van Kampen DA Willems WJ, van Beers LW, Castelein RM, Scholtes VA, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). J Orthop Surg Res. 2013; 8:40.
23. Stanberry B. Telemedicine: barriers and opportunities in the 21st century. J Intern Med. 2000;247(6):615-628.
24. Car J, Tan WS, Huang Z, Sloot P, Franklin BD. eHealth in the future of medications management: personalisation, monitoring and adherence. BMC Med. 2017;15(1):73.
25. Mars M, Scott RE. WhatsApp in clinical practice: a literature review. stud health technol ınform. 2016; 231:82-90.
26. Ozsoy I, Kodak MI, Karartı C, Ozsoy G, Erturk A, Kahraman T. Intra- and ınter-rater reproducibility of the face-to-face and tele-assessment of timed-up and go and 5-times sit-to-stand tests in patients with chronic obstructive pulmonary disease. Copd. 2022;19(1):125-132.
27. Ozsoy G, Aksoy K. Intra- and inter- rater reliability of the face-to-face assessment and tele-assessment of performance-based tests in older adults. Eur Geriatr Med. 2024;15(3):601-607.
28. Ozsoy I, Uz AL. Reliability of tele-assessment of five repetition sit to stand and timed up and go tests in patients with non-specific chronic low back pain. Discover Health Systems. 2024;3(1):34.
29. Erekdag A, Sener IN, Zengin Alpozgen A, Gunduz T, Eraksoy M, Kurtuncu M. The agreement between face-to-face and tele-assessment of balance tests in patients with multiple sclerosis. Mult Scler Relat Disord. 2024; 90:105766.
30. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthritis Cartilage. 2012;20(12):1548-1562.
31. Stratford PW. Estimating the standard error of measurement from reliability studies. Physiother Can. 2004;56(1):27-30.
32. Yuksel E, Kalkan S, Cekmece S, Unver B, Karatosun V. Assessing minimal detectable changes and test-retest reliability of the timed up and go test and the 2-minute walk test in patients with total knee arthroplasty. J Arthroplasty. 2017;32(2):426-430.
33. Unver B, Kalkan S, Yuksel E, Kahraman T, Karatosun V. Reliability of the 50-foot walk test and 30-sec chair stand test in total knee arthroplasty. Acta Ortop Bras. 2015;23(4):184-187.
34. Saraç DC, Unver B, Karatosun V. Validity and reliability of performance tests as balance measures in patients with total knee arthroplasty. Knee Surg Relat Res. 2022;34(1):11.