JETR

JOURNAL OF
EXERCISE THERAPY
AND REHABILITATION

Journal of Exercise Therapy and Rehabilitation. 2025;12(3):227-237. DOI: 10.15437/jetr.1689407

ORIGINAL ARTICLE

Reliability of tele-evaluation of Timed Up and Go, Single Leg Stand, and 30 s Sit to Stand Tests in patients with low back pain

Bel ağrısı olan hastalarda Zamanlı Kalk Yürü, Tek Ayak Üzerinde Durma ve 30 Saniye Otur Kalk Testlerinin tele-değerlendirmesinin güvenirliği

Muhammed Zahid UZ[1], Hayriye YILMAZ[2], Bilge KARA[3]

Abstract

Purpose: This study aimed to examine the reliability of the Timed Up and Go (TUG), Single Leg Standing (SLS), and 30-Second Sit-to-Stand (30STS) tests, both between different raters (inter-rater) and within the same rater (intra-rater), when administered through face-to-face and tele-evaluation methods in individuals with low back pain (LBP).
Methods: Fifty individuals diagnosed with LBP and meeting the inclusion criteria participated in the study. Detailed demographic characteristics, including age, sex, and body mass index (BMI), were recorded. Functional tests were conducted both in a traditional face-to-face setting and under synchronous (real-time) and asynchronous (recorded) tele-evaluation conditions.
Results: Inter-rater reliability between face-to-face and tele-evaluation methods was found to be very high (TUG: ICC=0.999; SLS: ICC=0.998; 30STS: ICC=0.996). Similarly, inter-rater reliability between two tele-evaluation sessions was also excellent (TUG: 0.997; SLS: 0.999; 30STS: 0.999). Intra-rater reliability, representing repeated measurements by the same rater, was also high in synchronous tele-evaluations, with ICC values of 0.997, 0.925, and 0.924 for TUG, SLS, and 30STS, respectively.
Conclusions: The TUG, SLS, and 30STS tests demonstrated high reliability in tele-evaluation applications among individuals with LBP. These findings indicate that the tests are valid, feasible, and clinically useful tools for standardized, safe, and remote evaluation of functional capacity in a home environment.
Keywords: Low back pain, Telemedicine, Telerehabilitation, Reliability.

Öz

Amaç: *Bu çalışmanın amacı, bel ağrısı olan bireylerde Zamanlı Kalk-Yürü (Timed Up and Go; TUG), Tek Ayak Üzerinde Durma (Single Leg Standing; SLS) ve 30 Saniye Otur-Kalk (30-Second Sit-to-Stand; 30STS) testlerinin, hem aynı değerlendirici hem de farklı değerlendiriciler arasında, yüz yüze ve tele-değerlendirme yöntemleriyle güvenirliğini incelemekti.*
Yöntem: *Çalışmaya, bel ağrısı tanısı almış ve dahil edilme kriterlerini karşılayan 50 birey katıldı. Katılımcıların yaş, cinsiyet ve beden kitle indeksi (BKİ) gibi ayrıntılı demografik verileri kaydedildi. Fonksiyonel testler hem geleneksel yüz yüze ortamda hem de senkron (eş zamanlı) ve asenkron (farklı zamanlı) tele-değerlendirme koşullarında uygulandı.*
Bulgular: *Yüz yüze ve tele-değerlendirme yöntemleri arasında farklı değerlendiriciler arası (inter-değerlendirici) güvenirlik çok yüksek bulundu (TUG: ICC=0,999; SLS: ICC=0,998; 30STS: ICC=0,996). İki tele-değerlendirme oturumu arasındaki farklı değerlendiriciler arası (inter-değerlendirici) güvenirlik de oldukça yüksekti (TUG: 0,997; SLS: 0,999; 30STS: 0,999). Aynı değerlendirici tarafından yapılan tekrar ölçümler arasındaki (intra-değerlendirici) güvenirlik, senkron tele-değerlendirmede TUG, SLS ve 30STS için sırasıyla 0,997, 0,925 ve 0,924 olarak bulundu.*
Sonuç: *TUG, SLS ve 30STS testleri, bel ağrısı olan bireylerde tele-değerlendirme uygulamaları açısından yüksek düzeyde güvenirlik sergilemiştir. Bu bulgular, söz konusu testlerin fonksiyonel kapasitenin ev ortamında uzaktan, güvenli ve standardize bir biçimde değerlendirilmesi için geçerli, uygulanabilir ve klinik olarak yararlı araçlar olduğunu göstermektedir.*
Anahtar kelimeler: *Bel ağrısı, Tele-tıp, Tele-rehabilitasyon, Güvenirlik.*

1.İzmir Kâtip Çelebi University, Vocational School of Health Services, Izmir, Türkiye
2:Health Sciences University, Izmir Bozyaka Training and Research Hospital, Department of Physical Therapy and Rehabilitation,Izmir, Türkiye
3: Dokuz Eylul University, Faculty of Physical Therapy and Rehabilitation, Izmir, Türkiye
Corresponding Author Muhammed Zahid Uz: muhammedzahiduz@gmail.com
ORCID IDs (order of authors): 0000-0002-7338-838X; 0000-0002-1151-7190; 0000-0003-4503-5074
Received: May 2, 2025. Accepted: November 29, 2025.

INTRODUCTION

Low back pain (LBP) is defined as discomfort or pain occurring between the lower rib margins and the gluteal folds, with or without radiating symptoms to the lower extremities. It represents one of the most prevalent musculoskeletal disorders globally and is a major cause of disability, work loss, and reduced quality of life.[1] According to the Global Burden of Disease Study, more than 80% of adults experience at least one episode of LBP during their lifetime, and recurrence rates are high. The persistent nature and economic consequences of LBP make it an important public health problem.[1-3]

Conventional management of LBP primarily involves physical therapy and rehabilitation. However, access to such services is often limited-particularly in rural or underserved areas-due to workforce shortages, mobility limitations, and economic barriers.[2] These challenges have stimulated the development of telerehabilitation, an emerging approach that enables healthcare providers to deliver physiotherapy and monitoring remotely through video conferencing and digital tools.[2-4] Telerehabilitation offers advantages such as increased accessibility, reduced travel costs, and continuity of care during situations like the COVID-19 pandemic.[5,6]

In physical rehabilitation, objective and reliable functional tests are essential for evaluating a patient's mobility, balance, and overall functional status.[4] Among the most widely used are the Timed Up and Go (TUG), Single Leg Stand (SLS), and 30-Second Sit-to-Stand (30STS) tests. The TUG assesses functional mobility and dynamic balance by measuring the time needed to stand up, walk 3 m, turn, and sit down again.[7,8] The SLS evaluates static balance and postural control,[9] while the 30STS measures lower-limb muscle strength and endurance.[10-12] These tests are practical, inexpensive, and easily applicable in both clinical and tele-based contexts.[13,14]

Although numerous studies have verified the reliability of these tests in traditional face-to-face evaluations, their reliability under tele-evaluation conditions remains limited.[13,15,16] Recently, Ozsoy and Uz demonstrated excellent tele-reliability of the TUG and five-repetition sit-to-stand tests in individuals with chronic non-specific low back pain.[17] Similarly, Bowman et al. confirmed the feasibility and safety of remote administration of the 30STS test,[18] and Karim et al. verified the practicality of tele-TUG in older adults.[19] These studies suggest that tele-evaluation can be a valid tool for functional performance measurement;[20,21] however, evidence among individuals with LBP is still scarce.

Therefore, this study aimed to investigate the intra-rater and inter-rater reliability of the TUG, SLS, and 30STS tests when administered via tele-evaluation in individuals with LBP. By comparing face-to-face and tele-based outcomes, this research sought to determine whether these tests could be safely and reliably integrated into telerehabilitation programs.[6,22,23]

METHODS

Study design and ethical considerations

This methodological reliability study was conducted at the Department of Physical Therapy and Rehabilitation, University of Health Sciences İzmir Bozyaka Training and Research Hospital. The research protocol was approved by the institutional ethics committee (Approval No: 2022/142; October 19, 2022) and complied with the ethical principles of the Declaration of Helsinki.[24] All participants signed informed consent forms prior to participation.

Participants

A total of 50 participants (33 female, 17 male) with LBP were recruited using simple random sampling. Randomization was generated by computer to minimize evaluator bias[38]. Participants were aged between 18 and 65 years (mean±SD = 52.4±7.1 years) and had experienced pain for at least three months.

Inclusion criteria

- Age between 18-65 years
- Body mass index (BMI) < 30 kg/m²
- Literate and able to follow instructions
- Basic smartphone skills sufficient for video calls
- Voluntary participation

Exclusion criteria

- History of spinal surgery or major orthopedic intervention

- Structural spinal pathologies (scoliosis, kyphosis, spondylolisthesis)
- Neurological, metabolic, vascular or psychiatric disorders affecting balance or mobility
- Uncontrolled cardiopulmonary conditions
- Inability to complete the test protocol or withdrawal from the study

Demographic and clinical data collected included age, sex, height, weight, BMI and dominant side.

Evaluation tools

Oswestry Disability Index (ODI)

"The Oswestry Disability Index (ODI) is used to determine the degree of functional limitation caused by LBP, providing a standardized measure of how the condition affects daily activities. The ODI assesses functional disability related to LBP across ten items (0-5 scale per item). Scores are converted to percentages (0-100%), with higher values indicating greater disability.[4,13]

Timed Up and Go (TUG)

This functional assessment evaluates the time required for an individual to rise from a standard chair, walk 3 meters at a usual walking speed, turn around, return to the starting point, and sit down again. Participants' performance was recorded in seconds using a stopwatch, and this measurement served as the primary outcome variable.[7] In line with established guidelines for conducting evaluations in the home environment, the appropriate chair dimensions for the TUG test were predetermined. These required measurements were communicated to all participants, and chairs that met the optimal specifications were identified based on the participants' feedback.

Single Leg Stand (SLS)

Participants stood on their preferred leg while crossing arms over the chest; time was recorded until the raised foot touched the floor or arms moved for balance.[9] Participants who achieved a score of less than 10 seconds on their first attempt were allowed to repeat the test. When necessary, they were given a rest period between trials to allow sufficient recovery. This test assessed static postural control.

30-Second Sit-to-Stand (30STS)

The 30STS test measured lower-limb strength and endurance.[10-12] Participants stood up and sat down as many times as possible within 30 seconds from a 43 cm-high chair placed against a wall for safety. Participants completed the test three times, and the highest number of repetitions among the trials was documented as the final score. When necessary, they were permitted to rest for more than 5 minutes between trials to allow adequate recovery.

Procedures

Each participant was evaluated under both face-to-face and tele-evaluation conditions to compare reliability between methods.[8-10] Tele-evaluations were conducted via *WhatsApp* video calls in two forms: *Synchronous* (real-time supervised session by Evaluator 1), Asynchronous (video recordings analysed later by Evaluator 2).

Evaluator 1 initially performed all evaluation face-to-face in the clinical environment.

The same evaluation procedures were subsequently repeated remotely and synchronously by Evaluator 1.

The video recordings generated during these remote evaluations were reviewed asynchronously by Evaluator 2.

During the retest phase, Evaluator 1 again conducted the evaluation via synchronous tele-evaluation.

The recordings from the retest session were also evaluated asynchronously by Evaluator 2.

Face-to-face evaluations were performed in the clinic by Evaluator 1. The evaluation procedure is visually outlined in Figure 1.

The smartphone camera was placed horizontally at hip height, 3 m from the participant, ensuring the entire movement remained visible during testing.[18,19] Each participant performed three trials per test; the mean of the trials was used for analysis. Rest periods of ≥ 5 minutes were provided between tests to prevent fatigue. A second evaluation session was performed 24-48 hours later to assess test-retest reliability.[25]

Statistical analysis

Data analysis was performed using IBM SPSS Statistics v25. Normality was checked using the Kolmogorov-Smirnov test and histograms.[26]

COnsensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines provide a qualitative

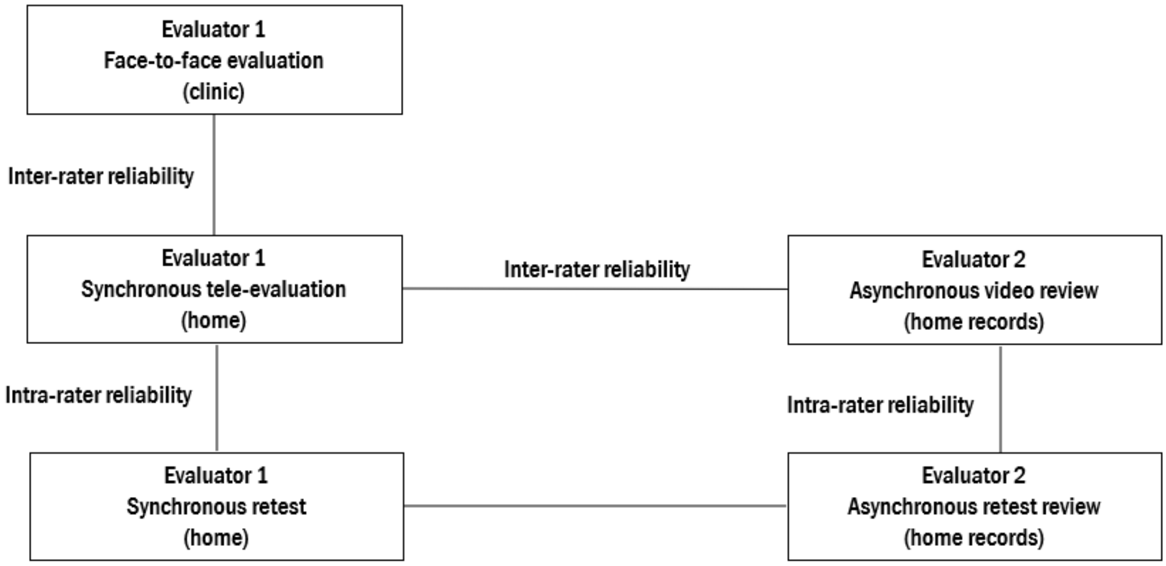

Figure 1. Evaluation procedure of the study.

rating framework in which a sample size of approximately 50 participants is categorized as 'adequate' for reliability studies.[24] Accordingly, in the present study, the selection of 50 participants was not based on a formal sample size calculation but was guided by COSMIN's indication of an acceptable minimum sample size for reliability research. This approach aligns with recommended methodological standards for conducting validity and reliability analyses.[24]

Reliability was evaluated using the Intraclass Correlation Coefficient (ICC) and Bland-Altman plots:[27]

- Intra-rater reliability: two-way mixed-effects model
- Inter-rater reliability: two-way random-effects model

Interpretation of ICC values was as follows: <0.50=poor, 0.50-0.75=moderate, 0.75-0.90=good, >0.90=excellent. The Standard Error of Measurement (SEM) and Smallest Detectable Change (SDC95%) were computed using established formulas.[26]

$SEM = SD * \sqrt{1-R}$

$SEM95\% = SEM*1.96$

$SDC95\% = SEM95\% * \sqrt{2}$

Systematic differences between sessions were examined using repeated-measures ANOVA.[38] A p-value<0.05 was considered statistically significant.

RESULTS

Sixty patients with a diagnosis of LBP who met the inclusion criteria were initially enrolled in the study. However, ten participants were excluded due to relocation or failure to complete the tele-evaluation process (Figure 2.) Consequently, the evaluation was completed for fifty patients. The mean age of the participants was 52.4±7.1 years, and 66% (n=33) were female. Detailed demographic characteristics of the participants are provided in Table 1.

Each participant successfully performed the Timed Up and Go (TUG), Single Leg Stand (SLS), and 30-Second Sit-to-Stand (30STS) tests, both through face-to-face clinical evaluation and tele-evaluations conducted in the home environment. The descriptive results of these functional evaluations are summarized in Table 2.

Inter-Rater Reliability

The study demonstrated high inter-rater reliability between the face-to-face and tele-evaluation methods for all three functional tests. The statistical models revealed exceptionally high levels of agreement between

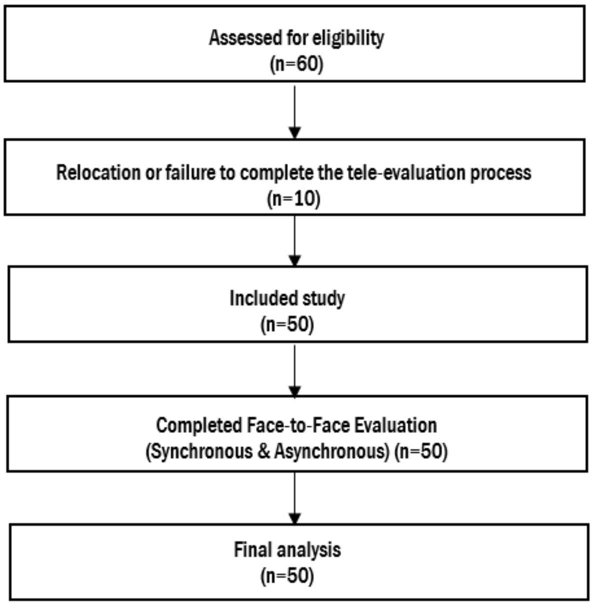

Figure 2. Flow chart of the study.

in-person and tele- evaluation results: TUG (ICC=0.999), SLS (ICC=0.998), and 30STS (ICC=0.996).

Evaluator experience was standardized to ensure consistency: the first tele-evaluator had 16 years of clinical physiotherapy experience, and the second had 14 years. Between these two tele-evaluators, the inter-rater reliability was also excellent: TUG (ICC=0.963), SLS (ICC=0.995), and 30STS (ICC=0.983).

Analysis of Bland-Altman plots (*Figures 3-5*) showed no systematic bias or trend, indicating the absence of proportional error or drift between in-person and tele- evaluation outcomes. A comprehensive summary of inter-rater reliability values, including ICC, 95% confidence intervals (CI), standard error of measurement (SEM), SEM 95%, and smallest detectable change (SDC 95%), is presented in Table 3.

Intra-Rater Reliability

The results from two-way mixed-effects models demonstrated excellent intra-rater reliability for synchronized tele-evaluations: TUG (ICC=0.997), SLS (ICC=0.925), and 30STS (ICC=0.924).

Similarly, the intra-rater reliability of asynchronous (video-based) tele-evaluation was excellent for the TUG (ICC=0.995) and good for both the SLS (ICC=0.807) and 30STS (ICC=0.839) tests. Visual inspection of Bland-Altman plots again showed no systematic differences, confirming measurement stability across sessions.

Repeated-measures ANOVA further indicated no significant differences between the two synchronized tele-evaluation sessions for the TUG (F=2.590, p=0.116), SLS (F=1.640, p=0.202), and 30STS (F=1.487, p=0.223) tests (Figures 3-5).

A detailed summary of the intra-rater reliability metrics for all three functional evaluations is provided in Table 3.

Overall, both synchronous and asynchronous tele-evaluation methods yielded reliability levels comparable to face-to-face evaluations for the TUG, SLS, and 30STS tests. The high intra- and inter-rater agreement supports the methodological robustness of tele-evaluation procedures in individuals with LBP. These results also highlight the consistency of test performance irrespective of evaluator experience and testing environment, aligning with the literature on remote functional evaluation reliability.

Table.1 Descriptive characteristics of patients.

	Mean±SD
Age (years)	52.4±7.1
Height (m)	1.66±0.08
Weight (kg)	81.6±13.2
Body mass index (kg/m²)	29.59±4.92
Oswestry Disability Index	54.6±9.1
	n (%)
Gender	
Male	17 (34)
Female	33 (66)
Educational status	
Illiterate	1 (2)
Primary-secondary education	37 (74)
High school	9 (18)
University	3 (6)
Social security	
Yes	48 (96)
No	2 (4)
Marital Status	
Married	48 (96)
Single	2 (4)
Dominant Side	
Right	44 (88)
Left	6 (12)

Table.2 Range of data of all evaluations of the tests.

	Face-to-face evaluation in clinic Evaluator 1	Tele- evaluation in home (synchronized) Evaluator 1	Retest tele-evaluation in home (synchronized) Evaluator 1	Evaluation in home from video records (asynchronized) Evaluator 2	Retest evaluation in home from video records (asynchronized) Evaluator 2
	Mean±SD	Mean±SD	Mean±SD	Mean±SD	Mean±SD
TUG (sec)	9.31±2.42	9.30±2.46	9.33±2,42	9.71±2.48	9.78±2.31
SLS (sec)	26.92±24.27	26.90±24.25	26.93±24.27	25.44±24.28	24.40±22.72
30STS (n)	9.78±2.36	9.74±2.30	9.78±2.36	9.70±1.98	9.58±2.13

n: number of repetitions. TUG: Timed Up and Go Test. SLS: Single Leg Stance Test. 30STS; 30-s Sit-to-Stand Test.

Table.3 Reliability results of the tests.

	ICC (%95CI)	SEM	SEM95%	SDC95%
Inter-reliability between face to face and tele- evaluation				
TUG (sec)	0.999 (0.998-0.999)	0.03	0.06	0.08
SLS (sec)	0.998 (0.997-0.999)	0.18	0.35	0.49
30STS (n)	0.996 (0.994-0.998)	0.03	0.06	0.08
Inter-reliability between two tele-evaluators				
TUG (sec)	0.999 (0.998-0.999)	0.002	0.004	0.006
SLS (sec)	0.999 (0.998-0.999)	0.003	0.006	0.008
30STS (n)	0.990 (0.974-0.996)	0.06	0.012	0.18
Intra rater-reliability of the tele-evaluation (synchronized)				
TUG (sec)	0.997 (0.996-0.998)	2.06	4.03	5.71
SLS (sec)	0.925 (0.863-0.958)	0.11	0.22	0.31
30STS (n)	0.924 (0.871-0.956)	0.11	0.22	0.31
Intra rater-reliability of the tele-evaluation from video records (asynchronized)				
TUG (sec)	0.995 (0.992-0.997)	2.07	4.05	5.73
SLS (sec)	0.807 (0.721-0.908)	0.19	0.37	0.52
30STS (n)	0.839 (0.733-0.905)	0.17	0.33	0.47

ICC: intraclass correlation coefficient. CI: confidence interval. SEM: Standard Error of Measurement (with a 95% confidence interval). SDC: Smallest Detectable Change (with a 95% confidence interval). TUG: Timed Up and Go Test. SLS: Single Leg Stance Test. 30STS:30-s Sit-to-Stand Test. n: number of repetitions.

DISCUSSION

This study evaluated the intra- and inter-rater reliability of three widely used functional tests—TUG, SLS, and 30STS—when performed via tele-evaluation in individuals with LBP. The results revealed good-to-excellent reliability for all tests, confirming that tele-evaluation is a feasible and accurate method for assessing physical performance in this population.[17-19,20,21]

The present findings align with previous research showing that remote administration of functional tests yields consistent and reproducible results. Ozsoy and Uz found excellent reliability for the tele-assessed TUG and sit-to-stand tests in chronic LBP patients.[17] Similarly, Bowman et al. demonstrated that the 30STS test can be safely administered through telehealth platforms,[18] and Karim et al. reported comparable findings for the TUG test in older adults.[19] Together, these studies support the validity of remote performance testing in musculoskeletal and geriatric populations,[22] further reinforcing the consistency observed in the present study.

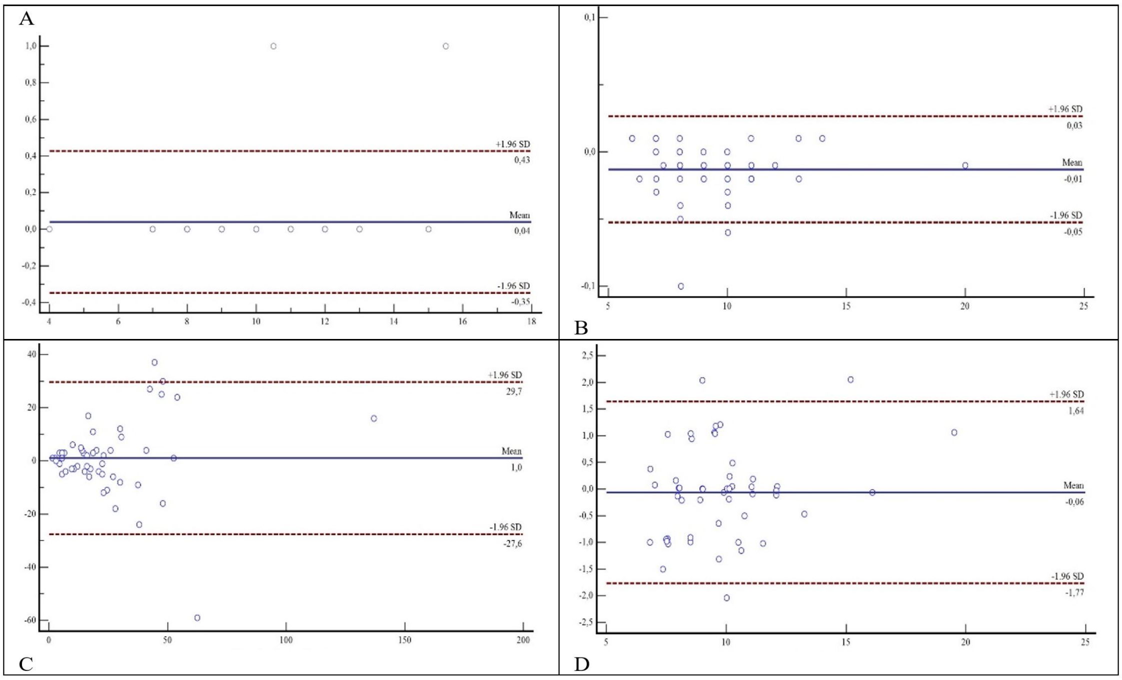

Figure 3. Bland-Altman plots for the Timed and Go Test. The means on the x axis are the average of two trials for the Timed and Go Test, and the differences between Timed and Go Test scores are in the y axis. (A) Inter-reliability between face-to-face and tele-evaluation. (B) Inter-reliability between two tele-raters. (C) Intra-rater reliability of the tele-evaluation (synchronized). (D) Intra-rater reliability of the tele-evaluation from videorecords (asynchronized). The 95% limits of agreement are depicted (dashed line). SD, standard deviation.

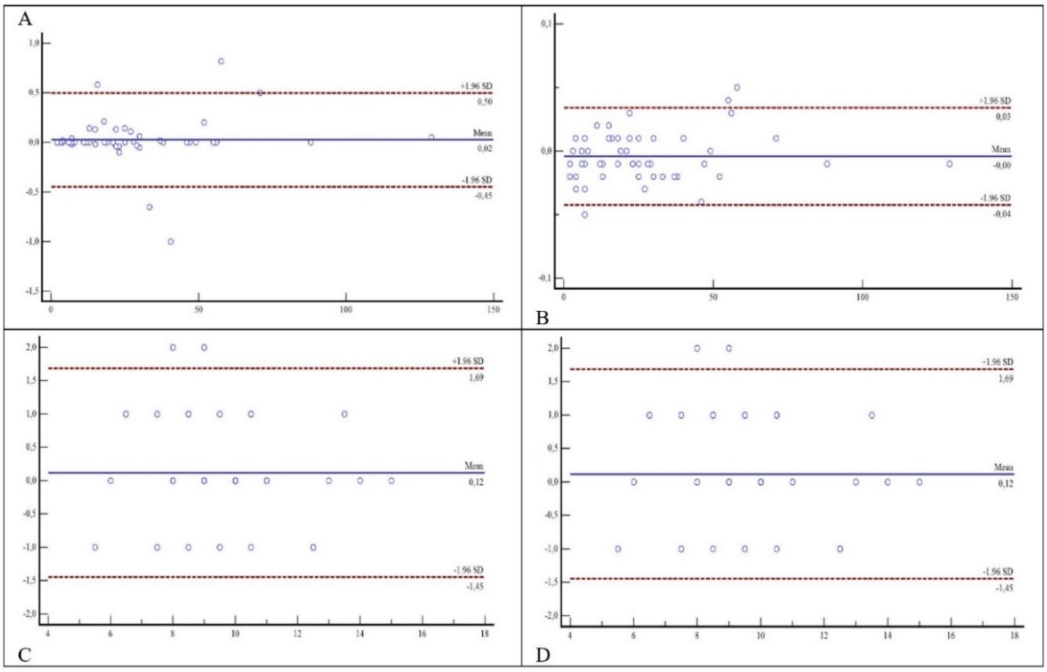

Figure 4. Bland-Altman plots for the Single Leg Stance Test. The means on the x axis are the average of two trials for the Single Leg Stance Test, and the differences between Single Leg Stance Test scores are in the y axis. (A) Inter-reliability between face-to-face and tele-evaluation. (B) Inter-reliability between two tele-raters. (C) Intra-rater reliability of the tele-evaluation (synchronized). (D) Intra-rater reliability of the tele-evaluation from videorecords (asynchronized). The 95% limits of agreement are depicted (dashed line). SD, standard deviation.

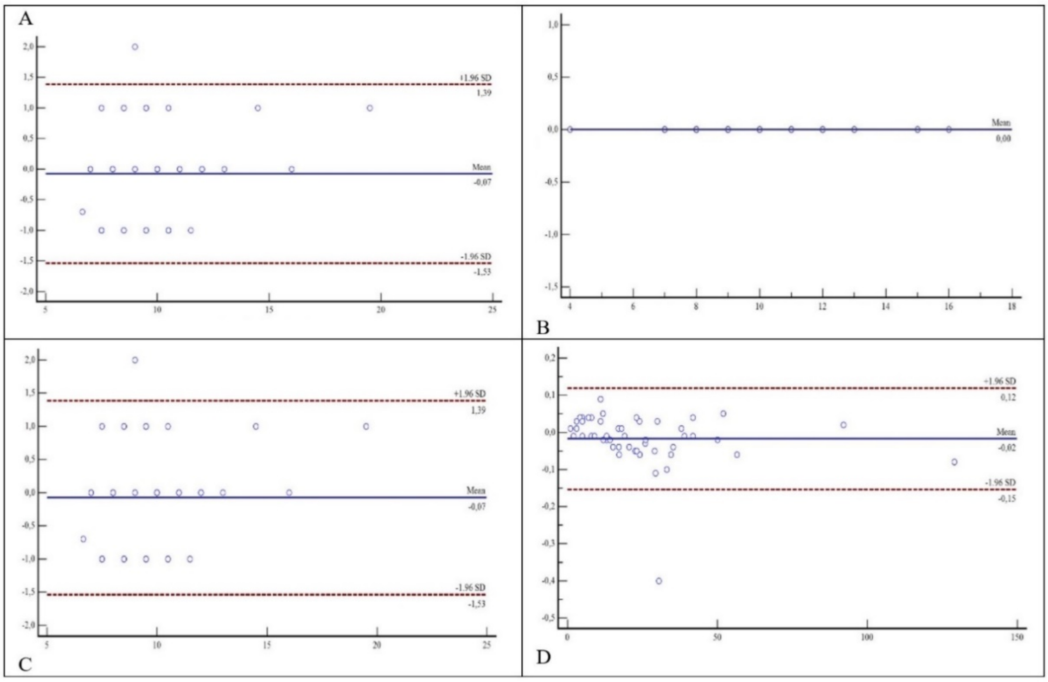

Figure 5. Bland-Altman plots for the 30-s Sit-to-Stand Test. The means on the x axis are the average of two trials for the 30-s Sit-to-Stand Test, and the differences between 30-s Sit-to-Stand Test scores are in the y axis. (A) Inter-reliability between face-to-face and tele-evaluation. (B) Inter-reliability between two tele-raters. (C) Intra-rater reliability of the tele-evaluation (synchronized). (D) Intra-rater reliability of the tele-evaluation from videorecords (asynchronized). The 95% limits of agreement are depicted (dashed line). SD, standard deviation.

The relatively lower ICC values observed for the SLS test appear to be associated with limitations in camera angles during tele-evaluation and the difficulty of capturing subtle postural adjustments that occur during balance tasks.[9] A similar pattern was noted for the 30STS test, particularly in asynchronous evaluations, where challenges in accurately identifying the start and end points of the movement from video recordings, as well as issues such as reduced resolution and frame drops caused by rapid repetitions, may diminish intra-rater consistency. These findings are in line with previous reports indicating that both static and dynamic balance evaluations are sensitive to visual perspective, camera distance, and environmental factors in tele-evaluation settings.[23,29] Despite these constraints, the acceptable reliability demonstrated by both SLS and 30STS tests in synchronous and asynchronous tele-evaluation formats suggests that, when appropriate camera positioning, adequate lighting, and standardized recording procedures are ensured, these tests remain feasible tools for remote functional evaluation.

Tele-evaluation offers several clinical advantages, such as reduced transportation barriers, improved accessibility for individuals in rural areas, and enhanced patient engagement in home environments.[2,5,6] Furthermore, the reliability demonstrated here suggests that remote evaluation may complement traditional in-person physiotherapy, particularly in follow-up and maintenance phases where continuity of care and convenience are critical.[20,22]

The excellent reliability of TUG and 30STS may stem from their clearly defined start-end actions and high visual detectability of gross motor movements, which minimize interpretation errors.[7,11,12] Conversely, the SLS test requires precise visual feedback to detect subtle sway patterns, explaining its slightly reduced ICC values in asynchronous conditions.[9] These differences highlight the importance of selecting appropriate camera perspectives and ensuring sufficient video quality when administering balance-focused tests remotely.

The absolute reliability of measurements and tests depends primarily on the precision of repeated test results. SEM and SDC95% are more clinically useful than ICC values because they are expressed in the same units as the instruments. In clinical practice, SDC95% can aid in categorizing research participants as either 'changed' or 'unchanged'. For instance, in the context of synchronous tele-SLS, SEM95% was determined to be 0.22 seconds. This indicates that if a patient's SLS score is 10 seconds, the actual score, 95% of the time, will fall within the range of 9.78 to 10.22 seconds. Under the same test conditions, SDC95% was calculated to be 0.31 seconds. Consequently, for a patient with an SLS score of 10 seconds, any score between 9.69 and 10.31 seconds would be considered within the margin of error.[26]

In repeated testing, it is anticipated that 95% of the time, the performance will accurately reflect an 'unchanged' status. However, it is essential to emphasize that these values represent measurement error rather than minimum clinically important difference values. According to the findings from repeated-measures ANOVA analysis, the results of the TUG, SLS, and 30STS tests did not exhibit statistically significant systematic errors, indicating the absence of consistent discrepancies between test trials.[30]

This study also contributes methodologically by providing explicit details on evaluator training, randomization, and camera setup—factors that have been underreported in previous tele-evaluation research.[20,22,26] Additionally, the participants' BMI, which was close to 30 kg/m², might have influenced test performance, as overweight individuals often demonstrate reduced balance and endurance capacity.[1,31]

Consistent with prior systematic reviews, our results reaffirm the feasibility and safety of telerehabilitation in musculoskeletal conditions, while underscoring the importance of standardizing video protocols and environmental parameters.[5,22,28] These findings align with growing evidence that tele-evaluation can deliver outcomes comparable to in-person evaluations,[6,23,32,33] provided that technical and procedural standards are carefully maintained.

Finally, this study adds to the expanding body of literature supporting digital physiotherapy, which has become increasingly relevant in the post-COVID-19 era.[34,35] Ensuring patient privacy, adequate technology access, and user-friendly telehealth interfaces will be essential for the widespread and sustainable implementation of such approaches.[36]

Limitations

This study has several limitations. First, although the sample size (n=50) aligns with COSMIN's recommendations for reliability studies, a larger and more heterogeneous population would enhance the generalizability of the findings. Second, the participants were recruited from a single rehabilitation center, limiting external validity. Third, the mean BMI was close to 30 kg/m², and overweight status might have influenced postural control and endurance. Fourth, tele-evaluation accuracy depends on technical factors such as camera placement, lighting, and internet quality, which may vary across home environments. Fifth, all tests were performed under the supervision of a single research team, which may restrict inter-center comparability. Future studies should include multiple centers and assess test-retest reliability across different time points and devices to enhance generalizability and external validity.

Conclusions

The current study demonstrated that the Timed Up and Go (TUG), Single Leg Stand (SLS), and 30-Second Sit-to-Stand (30STS) tests exhibit good-to-excellent intra- and inter-rater reliability when administered via tele-evaluation in individuals with low back pain. These findings confirm that remote administration of functional performance tests is both feasible and reliable, suggesting that such evaluations may be safely incorporated into telerehabilitation programs.

Tele-evaluation offers significant potential for improving accessibility and continuity of rehabilitation services, especially in rural regions and post-pandemic healthcare settings. Further large-scale studies are warranted to standardize tele-evaluation protocols and determine the long-term clinical utility of digital rehabilitation systems.

Acknowledgement: We thank the participants for their consent and cooperation.

Authors' Contributions: **MZU:** Conception, study design, and manuscript revision, analysis, validation, interpretation of data for the work, and writing original draft, approved the final version of the manuscript; **HY:** Conception, study design, and manuscript revision,investigation, and methodology, approved the final version of the manuscript; **BK:** Conception, study design, and manuscript revision, supervision, writing-review/editing, approved the final version of the manuscript.

Funding: *None*

Conflicts of Interest: *None*

Ethical Approval: The study was approved by the National Institutional Review Board and Ethics Committee of the University of Health Sciences, Izmir Bozyaka Training and Research Hospital (approval number 2022/142).

REFERENCES

1. Andersson GBJ. Epidemiological features of chronic low-back pain. Lancet. 1999;354:581-585.
2. Truter P, Russell T, Fisher A. The feasibility of a novel telerehabilitation system for orthopaedic rehabilitation. Physiotherapy. 2014;100:233-238.
3. Dario AB, Cabral AM, Macedo LC, et al. Effectiveness of telerehabilitation on pain and disability patients with chronic low-back pain: A systematic review and meta-analysis. Eur J Pain. 2017;21:1880-1889.
4. Palacín-Marín F, Esteban-Morenilla J, Carrasco MJ, et al. Reliability of clinical tests for lumbar instability and motor control in patients with chronic low back pain. Eur Spine J. 2013;22:305-312.
5. Chen J, Jin W, Zhang XX, et al. Telerehabilitation approaches for stroke patients: Systematic review and meta-analysis. Stroke. 2015;46:347-352.
6. Russell T, Buttrum P, Wootton R, Jull G, et al. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty: A randomized controlled trial. J Bone Joint Surg Am. 2011;93:113-120.
7. Simmonds MJ, Olshen RA, Bove EA, et al. Assessment of functional mobility in low-back-pain patients: Timed Up and Go test reliability. Spine. 1998;23:1155-1160.
8. Podsiadlo D, Richardson S. The Timed Up and Go: A test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39:142-148.
9. Lin SI, Woollacott MH. Postural control and the influence of somatosensory input in older adults: Functional performance implications. J Gerontol A Biol Sci Med Sci. 2002;57:B798-B804.
10. 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.
11. Bohannon RW. Reference values for the five-repetition sit-to-stand test: A descriptive meta-analysis. Phys Ther. 2006;86:1128-1135.
12. Kwan MMS, Lin SI, Chen CH, et al. Validation of the 30-second chair-stand test in community-dwelling older people. Age Ageing. 2011;40:558-563.
13. Palazón-Bru A, Tomás-Rodríguez MI, Gil-Guillén VF, et al. Reliability of physical performance tests in low back pain patients. Eur Spine J. 2017;26:97-103.
14. Denehy L, Berney S, Elliott D. Telehealth delivery of physical performance tests: A systematic review. J Telemed Telecare. 2019;25:397-407.
15. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155-163.
16. Stratford PW, Kennedy DM, Pagura SM. The reliability of the 30-second chair-stand test in patients with hip and knee osteoarthritis. Physiother Can. 2003;55:23-29.
17. 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. Discov Health Syst. 2024;3:34.
18. Bowman A, Denehy L, Benjemaa A, et al. Feasibility and safety of the 30-second sit-to-stand test delivered via telehealth: An observational study. PM R. 2023;15:31-40.
19. Karim R, Jasper A, Uy SJM, et al. Feasibility of Timed Up and Go administered via telerehabilitation in older adults: A pilot study. Top Geriatr Rehabil. 2023;39:229-234.
20. Tousignant M, Boissy P, Moffet H, et al. Patients' satisfaction with in-home telerehabilitation after total knee arthroplasty: A randomized controlled trial. Telemed J E Health. 2011;17:422-426.
21. Cottrell MA, Hill AJ, O'Leary SP, et al. Clinical outcomes for patients receiving telehealth-delivered physical therapy: A systematic review and meta-analysis. Telemed J E Health. 2018;24:628-635.

22. Suso-Martí L, La Touche R, Herranz-Gómez A, et al. Effectiveness of telerehabilitation in musculoskeletal disorders: Systematic review and meta-analysis. Phys Ther. 2020;100:1906-1921.
23. Cottrell MA, Russell TG. Telehealth for musculoskeletal physiotherapy: Making virtual physical therapy a reality. Phys Ther. 2020;100:830-842.
24. de Vet HCW, Terwee CB, Mokkink LB, et al. Measurement in Medicine: A Practical Guide. Cambridge (UK): Cambridge University Press; 2011.
25. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull. 1979;86:420-428.
26. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice (3rd ed.). Philadelphia (PA): F.A. Davis; 2015.
27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;327:307-310.
28. Cottrell MA, Galea OA, O'Leary SP, et al. Real-time telerehabilitation for the treatment of musculoskeletal conditions is effective and comparable to standard practice: A systematic review and meta-analysis. Clin Rehabil. 2017;31:625-638.
29. W, Green H, Sikhosana N, et al. Effectiveness of telehealth cardiac rehabilitation programs on health outcomes of patients with coronary heart diseases: An umbrella review. J Cardiopulm Rehabil Prev. 2024;44:15-25.
30. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19:231-240.
31. Guralnik JM, Ferrucci L, Simonsick EM, et al. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. N Engl J Med. 1995;332:556-561.
32. Moffet H, Tousignant M, Nadeau S, et al. In-home telerehabilitation compared with face-to-face rehabilitation after total knee arthroplasty: A non-inferiority randomized controlled trial. J Bone Joint Surg Am. 2015;97:1129-1141.
33. Pastora-Bernal JM, Martín-Valero R, Barón-López FJ, et al. Evidence of benefit of telerehabilitation after orthopedic surgery: A systematic review. J Med Internet Res. 2017;19:e142.
34. Bashshur RL, Shannon GW, Smith BR, et al. The empirical foundations of telemedicine interventions for chronic disease management. Telemed J E Health. 2014;20:769-800.
35. Cottrell MA, O'Leary SP, Hill AJ, et al. Service provision via telerehabilitation for patients with musculoskeletal conditions: A clinical commentary. J Orthop Sports Phys Ther. 2019;49:736-739.
36. Chen J, Cheng H, Wang Y. Factors influencing patients' acceptance of telemedicine services: An integrative review. Int J Environ Res Public Health. 2019;16:4223.