ORIGINAL SCIENTIFIC PAPER

Surface Effects on the Enhanced Paper Grip Test Results in Healthy Adults

Nenad Nedović[1,2], Milica Vidović[1], Dragana Vukov[1], Danilo Vujičić[1], Stevan Jovanović[1,3], Panagiotis Chatzistergos[4], Žiga Kozinc[5,6]

Affiliations: [1]Academy of Applied Studies Belgrade, College of Health Sciences, Serbia, [2]University of Belgrade, Institute for Medical Research, Belgrade, Serbia, [3]University of Kragujevac, Faculty of Medical Sciences, Kragujevac, Serbia, [4]School of Science and Engineering, University of Dundee, Dundee, UK, [5]University of Primorska, Faculty of Health Sciences, Izola, Slovenia, [6]Ludwig Boltzmann Institute for Rehabilitation Research, Vienna, Austria

Correspondence: Ž. Kozinc. University of Primorska, Faculty of Health Sciences, Polje 42, SI-6310 Izola, Slovenia. E-mail: ziga.kozinc@fvz.upr.si

Abstract

The enhanced paper grip test (EPGT) offers a practical approach to foot-strength assessment, but variation in testing surfaces may affect comparability across field settings. This study examined surface effects and within-session reliability in 20 healthy adults (20–68 years). Participants completed three trials per leg on paper, cardboard and linoleum. Surface comparisons used the mean of three peak-force measurements. Attachment–material friction was characterized separately under controlled loading. Repeated-measures ANOVA with Greenhouse–Geisser correction detected no surface effect (right: p = 0.384; left: p = 0.221). All descriptive 95% confidence intervals for paired mean differences included zero. Single-trial consistency reliability was excellent (ICC(3,1) = 0.955–0.980), increasing for three-trial means (ICC(3,3) = 0.985–0.993). Single-trial typical error was 3.45–6.01 N (CV = 7.8–12.5%). Surface correlations were strong to very strong (r = 0.789–0.936), but Bland–Altman limits indicated variable individual agreement; the widest limits were −36.77 to 40.71 N. Skewed differences limit interpretation of these normal-theory limits. A sample-size sensitivity analysis indicated 80% power for a two-sided paired effect of dz = 0.660. No systematic surface-dependent differences were detected, but small effects and individual discrepancies cannot be excluded. Neither surface interchangeability nor a negligible biological role of friction was established.

Keywords: *enhanced paper grip test, toe flexor strength, surface comparability*

@MJSSMontenegro
SURFACE EFFECTS ON ENHANCED PAPER GRIP TEST
http://mjssm.me/?sekcija=article&artid=332

Cite this article: Nedović, N., Vidović, M., Vukov, D., Vujičić, D., Jovanović, S., Chatzistergos, P., Kozinc, Ž. (2027) Surface Effects on the Enhanced Paper Grip Test Results in Healthy Adults. *Montenegrin Journal of Sports Science and Medicine,* 23 (1), Ahead of print. https://doi.org/10.26773/mjssm.270307

Introduction

Toe and foot flexor muscles contribute to arch support, foot stabilization and force transmission during propulsion (Yamauchi & Koyama, 2019). These functions provide a biomechanical rationale for examining their contribution to sporting tasks. In high-level athletes, metatarsophalangeal joint flexion torque has been associated with maximal-speed sprinting kinetics (Tourillon et al., 2024). Beyond sport, foot muscle weakness has been linked to an increased risk of falling (Mansi et al., 2023). The efficacy of foot strength training in improving posture, balance, and self-reported function further confirms the clinical significance of these

Received: 07 March 2026 | Accepted after revision: 25 October 2026 | Early access publication date: 01 November 2026 | Final publication date: 15 March 2027

© 2026 by the author(s). License MSA, Podgorica, Montenegro. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY).

Conflict of interest: None declared.

muscle groups (Jaffri et al., 2023). However, there remains a significant gap between the recognized importance of foot muscle strength for performance and its practical evaluation in clinical settings. The literature describes several reliable instruments for laboratory-based measurement, including pressure platforms (Mickle et al., 2016) and custom-made dynamometers (Man et al., 2016). However, the reliance on specialized technology remains a major obstacle, preventing these methods from becoming a routine part of clinical assessment.

The Enhanced Paper Grip Test (EPGT) (Chatzistergos et al., 2020) has recently emerged as a practical, cost-effective method for assessing foot-related force production. During the test, the participant presses the hallux onto an attachment while the examiner pulls it horizontally with a dynamometer; peak force before release is recorded. The interaction between toe loading, the attachment and the supporting surface makes friction a plausible influence on the measured force. Several studies have reported good to excellent reliability and practical applicability of EPGT in diverse demographic groups, including various age cohorts (Mansi et al., 2024; Skuk et al., 2024; Tsekoura et al., 2023). In soccer and volleyball players, the EPGT has shown good-to-excellent reliability and can quantify inter-limb differences, supporting its potential use in athlete profiling and field-based screening (Skuk et al., 2024). In addition, findings from Žura et al. (2025) support the utility of the EPGT for identifying balance insufficiencies in people with multiple sclerosis. Collectively, these findings position the EPGT as a promising tool that bridges the gap between laboratory-based measurements and clinically feasible assessment of foot muscle strength in different populations.

Although the EPGT can be administered on various surfaces, such as cardboard or linoleum, it remains unclear to what extent surface properties influence the measured force output or the reliability of the test. Friction, in particular, is a key mechanical factor that may alter the interaction between the toes and the testing surface (Westling & Johansson, 1984; Andre et al., 2009); however, its potential impact on EPGT outcomes has not been systematically investigated. The relevant literature recommends conducting the EPGT on a firm surface, with a sheet of standard matte paper (e.g., A4) taped to the floor and positioned between the foot and the floor (Mansi et al., 2023). However, the necessity and effectiveness of this approach have not been tested. The issue is particularly relevant to field assessment, where testing may occur in different gyms, laboratories, training facilities and competition environments. Surface-dependent differences could compromise between-session and between-location comparability. Previous applications in soccer and volleyball players (Skuk et al., 2024) and track-and-field athletes (Sagadin et al., 2025) therefore motivate evaluating surface sensitivity, while evidence of within-session reliability alone cannot establish comparability across locations.

Therefore, the primary aim of this study was to compare the outcomes of the EPGT across different surface conditions and to determine the reliability of the measurements. Specifically, this study seeks to examine (i) the presence of systematic bias across repetitions, (ii) the relative and absolute reliability of the test on different surfaces, (iii) the correlations and agreement among surfaces, and (iv) potential differences in mean force values between surfaces.

Methods

Participants

The study cohort consisted of healthy adult volunteers, recruited via a convenience sampling approach that utilized social media platforms, direct email invitations, and personal networks. For the total sample (N = 20), the mean ± SD age, height and body mass were 47.6 ± 16.7 years (range: 20–68 years), 171.9 ± 9.9 cm, and 78.9 ± 15.4 kg respectively. Females (n = 13) were 49.3 ± 17.4 years old, 166.6 ± 6.6 cm tall, and weighed 74.0 ± 12.4 kg, whereas males (n = 7) were 44.3 ± 16.2 years old, 181.7 ± 7.0 cm tall, and weighed 87.9 ± 17.4 kg. Individuals were excluded from participation if they reported any current pain in the lower extremities, exhibited limitations in the active range of motion of the foot (as determined by a visual assessment conducted by a physical therapist), or had a history of significant lower limb injuries. Prior to their involvement in the research, all participants provided written informed consent. The study protocol received formal approval by the University of Primorska's Commission for Ethics in Human Subjects Research (approval number: 4264-19-6/23).

Study design

This was a repeated-measures study conducted in a single laboratory session. Participants performed the EPGT under three surface conditions: standard office paper, cardboard, and linoleum (Figure 1). The order of surface testing was randomized to minimize order effects. Before the formal testing, participants performed one submaximal familiarization trial on each surface and leg to ensure proper task execution and reduce potential learning effects. Each participant completed three maximal trials with each leg on each surface, with 60-s rest intervals between repetitions to limit fatigue. Verbal encouragement was provided during all trials to ensure maximal effort. Static friction coefficients were also measured in a separate controlled mechanical experiment to characterize the attachment–material interfaces; these measurements were not intended to reproduce the complete biological EPGT interaction.

Enhanced paper grip test procedure

Three surface conditions were tested: standard office paper (A4), cardboard, and linoleum (Figure 1). These commonly available materials were selected to examine whether changing the supporting material affected the measured EPGT force. Paper grammage, cardboard thickness and type, manufacturers, and detailed linoleum specifications were not recorded in the available study documentation, limiting exact replication of the material interfaces.

Testing was performed with participants seated on a standard chair without armrests, with hips and knees flexed at approximately 90°, feet placed flat on the ground, and upper limbs relaxed alongside the trunk (Chatzistergos et al., 2020; Žura et al., 2025). A custom-designed 3D-printable attachment (available on Zenodo (ID: 10033725)), attached to a handheld dynamometer (EasyForce, Meloq, Sweden), was positioned under the hallux. Participants were instructed to press their toes downward to grip the attachment as strongly as possible while keeping the heel in full contact with the floor and avoiding forward sliding of the foot. The examiner gradually applied a horizontally pulling force to the attachment via the dynamometer until the attachment was completely removed from underneath the hallux. The peak force

DOI 10.26773/mjssm.270307

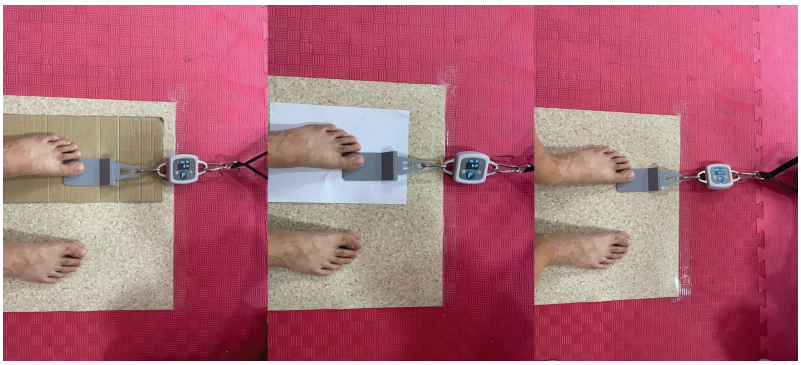


Figure 1. Measurement set-up with different surfaces.

(N) recorded at the moment of release was taken as the test outcome. Testing was repeated for both feet. All measurements were supervised and performed by the same examiner to minimize variability.

Static friction coefficients

Static friction coefficients were measured using an INSTRON 34SC-1 load frame equipped with a 50 N load cell (S-Beam, 2519-50N). A calibrated 1 kg weight was placed on the 3D-printed attachment, which was then pulled at a constant speed of 1 mm/min (Figure 2). Force data were recorded at 100 Hz, and each test was terminated immediately after a sudden drop in reaction force, indicating slippage. The maximum reaction force (F_{max}) prior to slippage was used for analysis. The static friction coefficient (μ_s) was calculated as the ratio of F_{max} to the total normal force (F_n) acting on the surface. F_n included the weight of both the 1 kg mass and the 3D-printed attachment (0.057 kg). Each test condition was repeated five times, and the reported friction coefficient represents the mean ± standard deviation of these replicates. Tests were conducted on three materials used in the study: the linoleum surface, a typical A4 paper, and typical cardboard.

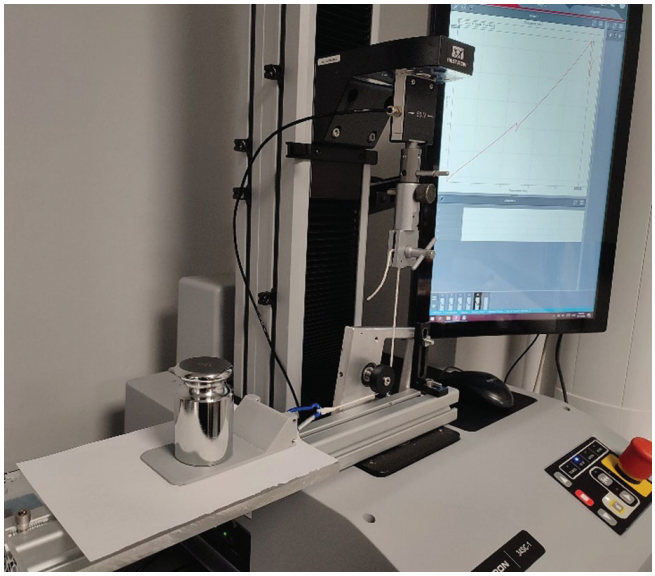


Figure 2. Testing set-up for the measurement of static friction coefficient.

Statistical analysis

All participant-level analyses used the complete dataset of 20 participants, with no missing trials or exclusions. Descriptive statistics were calculated for each repetition and each surface–leg condition. The participant score used for all surface ANOVAs, correlations and agreement analyses was the arithmetic mean of the three peak-force measurements for that surface and leg, rather than the maximum of the three trials. Raw force (N) was retained for reliability and principal inferential analyses. Body-mass-normalized scores ($N \cdot kg^{-1}$) were calculated individually by dividing each participant's three-trial mean by body mass and were summarized descriptively without additional hypothesis testing. Reliability of a specified measurement scale and descriptive characterization of strength are distinct purposes; normalization may alter between-participant variability and hence ICC estimates (Anders et al., 2025).

Systematic repetition effects and surface effects were examined using separate one-way repeated-measures ANOVAs for each relevant surface–leg condition and each leg, respectively. Greenhouse–Geisser correction was applied consistently to the degrees of freedom and p values, with partial eta squared (ηp^2) reported as the effect size. Surface inference was restricted to the omnibus tests. Paired surface mean differences and two-sided 95% CI were calculated as the mean difference ± $t_{0.975,19}$ × SD of differences/$\sqrt{20}$. These unadjusted intervals describe estimation uncertainty and were not used as separate post-hoc significance tests or evidence of equivalence. Repetition analyses were exploratory across six conditions, without multiplicity adjustment.

Within-session relative reliability was estimated using two-way mixed-effects consistency ICCs: ICC(3,1) for one trial and ICC(3,3) for the mean of three trials, with F-distri-

bution-based 95% CI. Participants were treated as random and the three ordered repetitions as fixed; the estimates pertain to the present examiner and protocol. The consistency definition excludes a common systematic repetition shift from the error term, so repetition effects were evaluated separately. ICC values above 0.90 were interpreted as excellent relative reliability (Koo & Li, 2016). Absolute random error was estimated from the residual mean square (MSE) of the participant-by-repetition model: $TE_1 = \sqrt{MSE}$ and $CV_1 = 100 \times TE_1/\text{grand mean}$. For the three-trial mean, model-based estimates were $TE_3 = TE_1/\sqrt{3}$ and $CV_3 = CV_1/\sqrt{3}$. These reductions assume independent trial errors of equal variance and do not remove systematic learning. TE confidence intervals used the chi-square distribution with 38 residual degrees of freedom; CV intervals were obtained by scaling the TE limits by the observed grand mean and are conditional approximations. These within-session estimates do not measure between-session or inter-examiner reproducibility.

Pearson correlations with Fisher-transformed 95% CI described preservation of rank order across surfaces; strong (0.70–0.89) and very strong (≥0.90) associations were distinguished (Akoglu, 2018). Individual-level agreement was quantified separately using Bland–Altman mean bias, SD of differences and conventional 95% limits of agreement (LoA = bias ± 1.96 × SD of differences) for each same-leg surface pair (Bland & Altman, 1999). The bias CI is identical to the descriptive CI of the paired mean difference. Difference distributions were checked using Shapiro–Wilk tests and numerical diagnostics, and the relationship between differences and paired averages was examined. Because several distributions were skewed and some differences varied with measurement magnitude, the normal-theory LoA were treated as descriptive summaries rather than validated 95% coverage limits. No acceptable agreement or equivalence margin had been specified. To assess the robustness of omnibus inference to non-normality, the surface ANOVAs were supplemented by 100,000 within-participant surface-label permutations using the same F statistic.

No a priori sample-size calculation was performed. A sensitivity analysis based on the noncentral t distribution indicated that N = 20 provides 80% power at two-sided α = 0.05 to detect a standardized paired mean difference of dz = 0.660, where dz is the mean difference divided by the SD of within-participant differences. This provides a benchmark for one paired contrast, not a power calculation for the three-level omnibus ANOVA or simultaneous comparisons. The sample was thus sensitive to moderate-to-large paired effects under the model assumptions, but small effects could not be excluded. No observed or retrospective power based on the obtained effect was calculated.

Results

Static friction coefficients

The results showed that the paper–3D printed attachment interface exhibited the highest static friction coefficient (0.433 ± 0.024), followed by linoleum (0.385 ± 0.032) and cardboard (0.308 ± 0.046).

Reliability

Descriptive force values, three-trial means and body-mass-normalized values are presented in Table 1. Right-leg force on cardboard increased from 46.2 ± 26.7 N in repetition 1 to 49.6 ± 27.3 N in repetition 3; the exploratory repetition effect remained significant after Greenhouse–Geisser correction ($F(1.86, 35.26) = 4.09$, $p = 0.028$, $\eta p^2 = 0.177$). No repetition effect was detected in the other conditions (corrected p = 0.070–0.250).

Table 1. Force descriptives and systematic repetition effects.

A. Peak force by repetition (N) and repetition ANOVA						
Surface and leg	Trial 1	Trial 2	Trial 3	F	p GG	ηp^2
Paper right	50.1 ± 29.2	52.6 ± 32.2	53.8 ± 33.8	2.57	0.102	0.119
Paper left	46.0 ± 25.0	48.4 ± 29.0	50.0 ± 30.8	2.19	0.140	0.103
Cardboard right	46.2 ± 26.7	47.8 ± 25.1	49.6 ± 27.3	4.09	0.028	0.177
Cardboard left	41.1 ± 22.3	43.4 ± 23.3	44.1 ± 24.4	3.02	0.070	0.137
Linoleum right	48.1 ± 27.4	51.1 ± 31.5	51.4 ± 28.7	2.10	0.149	0.100
Linoleum left	45.2 ± 20.0	43.4 ± 17.8	44.1 ± 18.2	1.44	0.250	0.071

B. Participant scores used for all surface comparisons			
Surface	Leg	Three-trial mean (N)	Three-trial mean ($N \cdot kg^{-1}$)
Paper	Right	52.2 ± 31.5	0.662 ± 0.347
Paper	Left	48.1 ± 27.9	0.612 ± 0.334
Cardboard	Right	47.9 ± 26.2	0.609 ± 0.294
Cardboard	Left	42.9 ± 23.1	0.540 ± 0.248
Linoleum	Right	50.2 ± 28.9	0.634 ± 0.294
Linoleum	Left	44.2 ± 18.5	0.560 ± 0.206

Values are mean ± SD; N = 20. Each trial contributes its peak force.

Single-trial consistency reliability was excellent across the six conditions (ICC(3,1) = 0.955–0.980), and reliability of the three-trial mean was higher (ICC(3,3) = 0.985–0.993). Single-trial TE ranged from 3.45 to 6.01 N and CV from 7.8% to 12.5%. For the three-trial mean, model-based TE ranged from 1.99 to 3.47 N and CV from 4.5% to 7.2% (Table 2). These random-error estimates exclude systematic repetition shifts.

DOI 10.26773/mjssm.270307

Table 2. Within-session reliability of single trials and three-trial means.

A. Relative reliability		
Surface and leg	ICC(3,1) (95% CI)	ICC(3,3) (95% CI)
Paper right	0.973 (0.945 to 0.989)	0.991 (0.981 to 0.996)
Paper left	0.955 (0.908 to 0.980)	0.985 (0.967 to 0.993)
Cardboard right	0.980 (0.958 to 0.991)	0.993 (0.986 to 0.997)
Cardboard left	0.971 (0.940 to 0.987)	0.990 (0.979 to 0.996)
Linoleum right	0.962 (0.923 to 0.984)	0.987 (0.973 to 0.995)
Linoleum left	0.966 (0.930 to 0.985)	0.988 (0.976 to 0.995)
B. Absolute random error in N		
Surface and leg	Single-trial TE (95% CI)	Mean-of-three TE (95% CI)
Paper right	5.19 (4.24 to 6.68)	2.99 (2.45 to 3.86)
Paper left	6.01 (4.91 to 7.74)	3.47 (2.84 to 4.47)
Cardboard right	3.76 (3.08 to 4.85)	2.17 (1.78 to 2.80)
Cardboard left	4.00 (3.27 to 5.15)	2.31 (1.89 to 2.97)
Linoleum right	5.67 (4.63 to 7.30)	3.27 (2.67 to 4.22)
Linoleum left	3.45 (2.82 to 4.45)	1.99 (1.63 to 2.57)
C. Absolute random error relative to the grand mean		
Surface and leg	Single-trial CV % (95% CI)	Mean-of-three CV % (95% CI)
Paper right	9.9 (8.1 to 12.8)	5.7 (4.7 to 7.4)
Paper left	12.5 (10.2 to 16.1)	7.2 (5.9 to 9.3)
Cardboard right	7.9 (6.4 to 10.1)	4.5 (3.7 to 5.9)
Cardboard left	9.3 (7.6 to 12.0)	5.4 (4.4 to 6.9)
Linoleum right	11.3 (9.2 to 14.5)	6.5 (5.3 to 8.4)
Linoleum left	7.8 (6.4 to 10.1)	4.5 (3.7 to 5.8)

N = 20, three trials. ICCs use a two-way mixed-effects consistency model. $TE_1 = \sqrt{MSE}$ with 38 residual degrees of freedom; $CV_1 = 100 \times TE_1$/grand mean. TE_3 and CV_3 divide the corresponding single-trial estimate and CI by $\sqrt{3}$.

Correlation and agreement among surfaces

Correlations between three-trial mean scores were strong to very strong (Table 3). For the right leg, r ranged from 0.789 (paper–linoleum) to 0.936 (paper–cardboard); for the left leg, r ranged from 0.841 (paper–linoleum) to 0.936 (cardboard–linoleum). These associations describe preservation of participant ranking, not agreement in absolute force.

Table 3. Correlations between surface scores.

Leg	Surface pair	Pearson r (95% CI)
Right	Paper–Cardboard	0.936 (0.842 to 0.975)
Right	Paper–Linoleum	0.789 (0.532 to 0.913)
Right	Cardboard–Linoleum	0.922 (0.810 to 0.969)
Left	Paper–Cardboard	0.848 (0.649 to 0.938)
Left	Paper–Linoleum	0.841 (0.634 to 0.935)
Left	Cardboard–Linoleum	0.936 (0.844 to 0.975)

N = 20. Scores are the mean of three trials per surface and leg. Confidence intervals use Fisher's z transformation.

The surface ANOVAs detected no systematic mean differences for the right leg ($F(1.21, 23.08) = 0.86$, $p = 0.384$, $\eta p^2 = 0.043$) or left leg ($F(1.48, 28.08) = 1.60$, $p = 0.221$, $\eta p^2 = 0.078$). Omnibus permutation checks also did not detect surface effects (right: $p = 0.502$; left: $p = 0.213$). Descriptive paired mean differences ranged from −2.35 to 5.22 N, and all 95% CI included zero (Table 4). The corrected left-leg cardboard–linoleum difference was −1.35 N (95% CI: −5.43 to 2.73 N).

Despite small group mean biases, individual differences between surfaces were variable. Conventional LoA were widest for right-leg paper–linoleum (−36.77 to 40.71 N) and narrowest for left-leg cardboard–linoleum (−18.44 to 15.74 N) (Table 4; Figure 3). All four comparisons involving paper showed departures from normality of differences (Shapiro–Wilk $p < 0.01$), and numerical diagnostics suggested magnitude-dependent differences in some comparisons. The LoA therefore require caution as descriptive estimates. Neither the correlations nor the absence of an omnibus surface effect establishes individual-level interchangeability.

Table 4. Omnibus surface comparisons and individual agreement.

A. Repeated-measures surface ANOVA				
Leg	F	GG degrees of freedom	p GG	ηp^2
Right	0.863	1.21, 23.08	0.384	0.043
Left	1.603	1.48, 28.08	0.221	0.078
B. Descriptive paired differences and Bland–Altman statistics (N)				
Leg and surface pair	Mean difference / bias (95% CI)	SD of differences	Conventional 95% LoA	
Right Paper–Cardboard	4.32 (−1.10 to 9.73)	11.58	−18.37 to 27.01	
Right Paper–Linoleum	1.97 (−7.28 to 11.22)	19.77	−36.77 to 40.71	
Right Cardboard–Linoleum	−2.35 (−7.58 to 2.88)	11.18	−24.27 to 19.57	
Left Paper–Cardboard	5.22 (−1.71 to 12.15)	14.80	−23.80 to 34.23	
Left Paper–Linoleum	3.87 (−3.59 to 11.32)	15.93	−27.35 to 35.08	
Left Cardboard–Linoleum	−1.35 (−5.43 to 2.73)	8.72	−18.44 to 15.74	

N = 20. All analyses use three-trial means. Differences are first surface minus second surface. Surface inference is restricted to panel A; panel B contains descriptive, unadjusted intervals and no post-hoc significance tests.

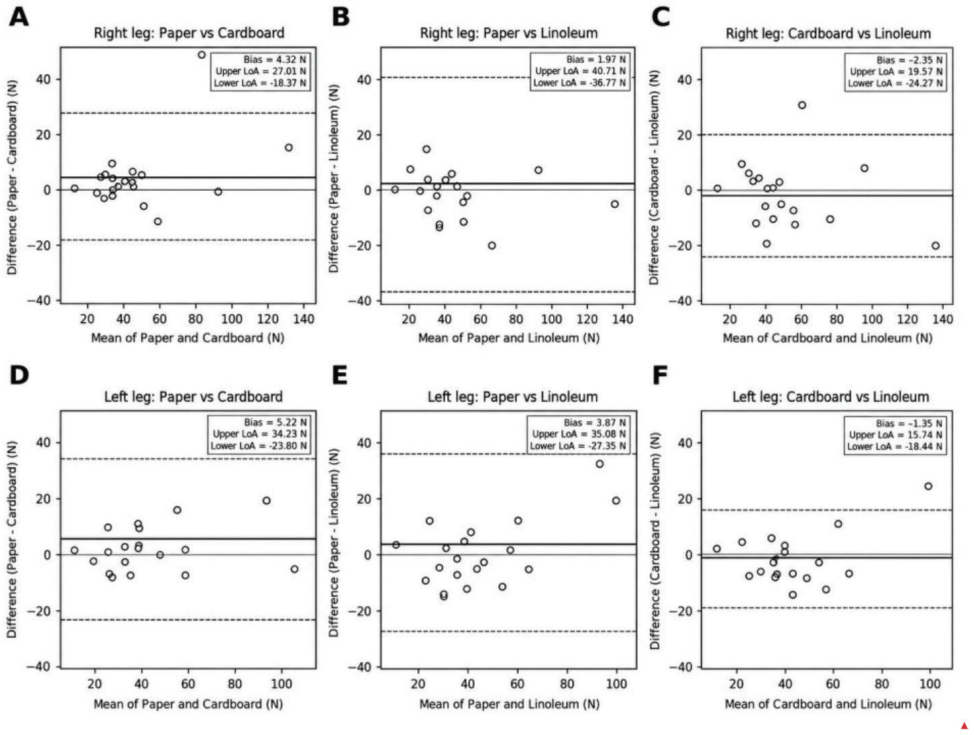


Figure 3. Bland–Altman plots showing agreement between EPGT force measurements obtained on different surfaces. Agreement between the three-trial mean EPGT scores obtained on paper, cardboard, and linoleum was examined separately for the right leg (A–C) and left leg (D–F). Panels show comparisons between paper and cardboard (A, D), paper and linoleum (B, E), and cardboard and linoleum (C, F).

Discussion

The primary aim of this study was to compare the outcomes of the EPGT across different surface conditions and to evaluate within-session reliability. No systematic surface-dependent differences in the mean of three EPGT force measurements were detected across paper, cardboard and linoleum. However, the confidence intervals permit non-zero effects, and the study's sensitivity benchmark was dz = 0.660 for a single paired contrast. The non-significant omnibus tests therefore do not establish equivalence or rule out small surface effects. This distinction matters in sport settings, where changing the available surface across laboratories, gyms, training facilities or competition venues may affect interpretation of repeated measurements.

Correlations were strong to very strong (r = 0.789–0.936), indicating substantial preservation of participant ranking across surfaces. In contrast, Bland–Altman analysis showed that individual absolute force values could differ appreciably, with conventional LoA as wide as −36.77 to 40.71 N for right-leg paper–linoleum. These findings distinguish three separate questions: the omnibus ANOVA concerns systematic group mean differences, correlation concerns relative ranking, and agreement concerns differences in an individual's absolute score. Moreover, skewed and potentially magnitude-dependent differences limit the coverage interpretation of constant LoA. No acceptable agreement

DOI 10.26773/mjssm.270307

margin was defined, and the results do not justify treating the tested surfaces as interchangeable for individual monitoring.

The mechanical experiment characterized the friction of the 3D-printed attachment against each material under controlled normal loading, with the highest reported coefficient for paper, followed by linoleum and cardboard. Friction is a property of a material pair and depends on the loading and contact conditions (Chowdhury et al., 2012). During biological EPGT testing, plantar skin and soft tissue, active hallux and toe loading, changing normal force, participant stabilization and manually applied pulling force also contribute to the interaction. Consequently, differences in isolated attachment–material friction coefficients coupled with an absence of detected mean EPGT differences cannot establish that friction is unimportant during the actual task. The friction experiment provides supportive mechanical characterization, rather than a mechanistic validation of EPGT or evidence that intrinsic toe-flexor strength is the primary determinant of its outcome.

Within-session relative reliability was excellent for individual trials (ICC(3,1) = 0.955–0.980), consistent with previous evidence of high relative reliability for dynamometry-based toe-strength assessment in children (Quinlan et al., 2021), adults (Ridge et al., 2017), and EPGT testing (Chatzistergos et al., 2020; Žura et al., 2025). However, high ICCs reflect discrimination between participants relative to random error and do not imply small absolute error. Single-trial TE was 3.45–6.01 N and CV was 7.8–12.5%. Averaging three trials increased ICCs to 0.985–0.993 and reduced model-based random error to TE = 1.99–3.47 N and CV = 4.5–7.2%, supporting a clearly specified repeated-trial scoring protocol. These estimates assume independent, equal-variance trial errors and do not establish between-session repeatability or between-surface agreement.

The right-leg cardboard repetition effect persisted after reanalysis, with force increasing across the three repetitions. Residual familiarization or learning may therefore have remained despite the preliminary submaximal familiarization trial. The finding arose in only one of six exploratory conditions and should not be interpreted as a general learning effect or a mechanism specific to cardboard. Consistency ICCs and residual TE do not incorporate a common shift across repetitions; averaging alone cannot remove such a shift. Future protocols should examine whether additional familiarization improves stability before formal measurements.

Raw-force reliability and normalized strength description address different purposes. Reliability was assessed in N to characterize the measurement scale used by the test, whereas individual body-mass normalization provides complementary descriptive information in a sample spanning 56–120 kg and including both sexes. Normalized three-trial mean scores ranged from 0.540 to 0.662 $N \cdot kg^{-1}$ across the six conditions. These descriptive values neither validate body mass as the optimal scaling variable nor establish reliability of the normalized score; no additional hypothesis tests on normalized data were performed.

The EPGT remains a practical, low-cost candidate for field assessment, but the present results support a cautious and standardized application. Previous studies have consistently demonstrated good to excellent reliability and feasibility of the EPGT in various populations, including older adults and athletes (Sagadin et al., 2025; Skuk et al., 2024; Tsekoura et al., 2023), while recent evidence indicates a meaningful association between EPGT performance, balance control, and functional independence (Mansi et al., 2023; Tsekoura et al., 2023; Žura et al., 2025). For sport science and sports medicine, these attributes support the EPGT as a candidate tool for foot-strength and inter-limb asymmetry assessment and as a complementary indicator of balance or lower-limb function. For longitudinal individual monitoring, the same surface, scoring rule and examiner technique should be retained where feasible, and small changes should be interpreted in relation to measurement error. The current study assessed only within-session reliability in healthy adults. Athlete-specific validity, reference values and responsiveness require separate evaluation before EPGT scores inform individual sport-related decisions.

Limitations

Several limitations should be acknowledged. The convenience sample was small and comprised healthy adults rather than athletes or clinical populations. Its broad age range (20–68 years) may have increased between-participant variability and hence ICCs, and the study was not designed to assess age- or sex-specific surface effects. The paired sensitivity benchmark indicates limited sensitivity to small effects and is not proof of adequate power for every surface contrast or for equivalence. Bland–Altman estimates are also imprecise with N = 20; skewed differences and possible dependence on force magnitude limit the generalizability of constant normal-theory LoA and model-based intervals. All observations were retained, including large between-surface discrepancies.

Pulling speed and angle were controlled manually. Variations in speed could influence the time available to develop or sustain toe loading and the force recorded at release, whereas deviations from horizontal pulling could alter the direction of loading and the normal contact force. These factors may produce within-examiner error even when all tests are performed by the same examiner. Their individual contributions were not measured, and the design cannot estimate inter-examiner reproducibility. In addition, incomplete material specifications restrict exact replication, and the separate friction experiment does not reproduce the full biological interface. Future work should examine standardized pulling procedures, material specifications, familiarization and between-session measurement error.

Conclusions

No systematic surface-dependent differences in the mean of three EPGT force measurements were detected across the three tested surfaces. Excellent within-session consistency reliability and strong to very strong surface correlations coexisted with appreciable random error and variable individual agreement. Averaging three trials reduced model-based random error, but the confidence intervals, sensitivity benchmark and Bland–Altman findings preclude claims of equivalence or individual-level interchangeability. For repeated individual assessment, consistent surfaces and procedures remain advisable. The controlled friction measurements characterize the tested attachment–material interfaces without establishing the contribution of friction or intrinsic toe-flexor strength during the biological task.

Declarations

Author contribution statement

N. Nedović: Conceptualization; Methodology; Data curation; Investigation; Formal analysis; Writing – original draft.
M. Vidović: Investigation; Data curation; Writing – review & editing.
D. Vukov: Investigation; Data curation; Writing – review & editing.
D. Vujičić: Methodology; Investigation; Resources; Writing – review & editing.
S. Jovanović: Methodology; Supervision; Writing – review & editing.
P. Chatzistergos: Methodology; Supervision; Writing – review & editing.
Ž. Kozinc: Conceptualization; Methodology; Formal analysis; Visualization; Writing – original draft; Writing – review & editing; Project administration.

Declaration of interests

The authors report there are no competing interests to declare.

Funding

The authors acknowledge the support by the Slovenian Research Agency through the research program KINSPO - Kinesiology for the effectiveness and prevention of musculoskeletal injuries in sports (P5-0443). The funder played no role in study conceptualization, data collection and preparation of the manuscript.

Ethics

All procedures conformed to the principles of the Declaration of Helsinki. The study procedures were approved by the University of Primorska's Commission for Ethics in Human Subjects Research (approval number: 4264-19-6/23).

Acknowledgements

None.

Data availability

Raw data pertaining to this study is freely available in Zenodo database (https://zenodo.org/records/17608124)

References

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Anders, C., Alex, C., Herzberg, M., & Mader, L. (2025). Normalized data may have limitations in determining the reliability of MVC measurements. *Scientific Reports*, *15*(1), 32930. https://doi.org/10.1038/s41598-025-20014-9

André, T., Lefèvre, P., & Thonnard, J.-L. (2009). A continuous measure of fingertip friction during precision grip. *Journal of Neuroscience Methods*, *179*(2), 224–229. https://doi.org/10.1016/j.jneumeth.2009.01.031

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. Statistical Methods in Medical Research, 8(2), 135–160. https://doi.org/10.1177/096228029900800204

Chatzistergos, P. E., Healy, A., Balasubramanian, G., Sundar, L., Ramachandran, A., & Chockalingam, N. (2020). Reliability and validity of an enhanced paper grip test; A simple clinical test for assessing lower limb strength. *Gait & Posture*, *81*, 120–125. https://doi.org/10.1016/j.gaitpost.2020.07.011

Chowdhury, M. A., Nuruzzaman, D. M., Mia, A. H., & Rahaman, M. L. (2012). Friction coefficient of different material pairs under different normal loads and sliding velocities. *Tribology in Industry, 34*(1), 18–23.

Jaffri, A. H., Koldenhoven, R., Saliba, S., & Hertel, J. (2023). Evidence for Intrinsic Foot Muscle Training in Improving Foot Function: A Systematic Review and Meta-Analysis. *Journal of Athletic Training*, *58*(11–12), 941–951. https://doi.org/10.4085/1062-6050-0162.22

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Man, H.-S., Leung, A. K.-L., Cheung, J. T.-M., & Sterzing, T. (2016). Reliability of metatarsophalangeal and ankle joint torque measurements by an innovative device. *Gait & Posture*, *48*, 189–193. https://doi.org/10.1016/j.gaitpost.2016.05.016

Mansi, M. K., Chockalingam, N., & Chatzistergos, P. E. (2023). An exploration of the mechanistic link between the enhanced paper grip test and the risk of falling. *The Foot*, *57*, 102059. https://doi.org/10.1016/j.foot.2023.102059

Mansi, M. K., Chockalingam, N., & Chatzistergos, P. E. (2024). The enhanced paper grip test can substantially improve community screening for the risk of falling. *Gait & Posture*, *108*, 157–163. https://doi.org/10.1016/j.gaitpost.2023.12.006

Mickle, K. J., Angin, S., Crofts, G., & Nester, C. J. (2016). Effects of Age on Strength and Morphology of Toe Flexor Muscles. *Journal of Orthopaedic & Sports Physical Therapy*, *46*(12), 1065–1070. https://doi.org/10.2519/jospt.2016.6597

Quinlan, S., Sinclair, P., Hunt, A., & Fong Yan, A. (2021). Excellent reliability of toe strength measurements in children aged ten to twelve years achieved with a novel fixed dynamometer. *Gait & Posture*, *85*, 20–24. https://doi.org/10.1016/j.gaitpost.2021.01.015

Ridge, S. T., Myrer, J. W., Olsen, M. T., Jurgensmeier, K., & Johnson, A. W. (2017). Reliability of doming and toe flexion testing to quantify foot muscle strength. *Journal of Foot and Ankle Research*, *10*(1). https://doi.org/10.1186/s13047-017-0237-y

Sagadin, U., Zupančič, M., & Kozinc, Ž. (2025). Application of enhanced paper grip test among track and field athletes: age difference, intra-session reliability and effects of fatiguing task. *Sportlogia*, *21*(1). https://doi.org/10.63356/spl.2025.008

Skuk, N., Chatzistergos, P. E., & Kozinc, Ž. (2024). Reliability of enhanced paper grip test for testing foot strength in volleyball and soccer players. *Physical Therapy in Sport*, *67*, 155–160. https://doi.org/10.1016/j.ptsp.2024.05.004

Tourillon, R., Michel, A., Fourchet, F., Edouard, P., & Morin, J.-B. (2024). Human foot muscle strength and its association with sprint acceleration, cutting and jumping performance, and kinetics in high-level athletes. *Journal of Sports Sciences, 42*(9), 814–824. https://doi.org/10.1080/02640414.2024.2367365

Tsekoura, M., Pantou, M., Fousekis, K., Billis, E., Gliatis, J., & Tsepis, E. (2023). Reliability and clinical applicability

 DOI 10.26773/mjssm.270307

of lower limp strength using an enhanced paper grip strength. *European Journal of Translational Myology*, *33*(4). https://doi.org/10.4081/ejtm.2023.11841

Westling, G., & Johansson, R. S. (1984). Factors influencing the force control during precision grip. *Experimental Brain Research*, *53*(2). https://doi.org/10.1007/BF00238156

Yamauchi, J., & Koyama, K. (2019). Force-generating capacity of the toe flexor muscles and dynamic function of the foot arch in upright standing. *Journal of Anatomy*, *234*(4), 515–522. https://doi.org/10.1111/joa.12937

Žura, E., Brecl Jakob, G., Chatzistergos, P., & Kozinc, Ž. (2025). Assessment of the enhanced paper grip test for quantifying balance deficits in people with multiple sclerosis: A reliability and applicability study. *Multiple Sclerosis and Related Disorders*, *95*, 106321. https://doi.org/10.1016/j.msard.2025.106321