

Article Classification:
Original article

1RM MEASUREMENT IN THE STANDING CALF RAISE: EVALUATION OF RELIABILITY AND MEASUREMENT ERROR

Sebastian Puschkasch-Möck¹, Michael Keiner²,
Klaus Wirth³, Andreas Konrad⁴, Konstantin Warneke⁵

OPEN ACCESS ¹ Olympic Training and Testing Center of Hessen, Frankfurt, Germany

Submitted: 11 April 2025

Accepted: 27 June 2025

Published: 14 July 2025

² German University of Health and Sport, Ismaning, Germany

³ University of Applied Sciences Wiener Neustadt, Wiener Neustadt, Austria,

⁴ University of Graz, Graz, Austria

⁵ University of Jena, Jena, Germany

ORCID 

Sebastian Puschkasch-Möck
<https://orcid.org/0000-0001-6277-2696>

Michael Keiner
<https://orcid.org/0000-0002-1817-1743>

Andreas Konrad
<https://orcid.org/0000-0002-5588-1824>

Konstantin Warneke
<https://orcid.org/0000-0003-4964-2867>

Cite this article as:
Sebastian Puschkasch-Möck, S.,
Keiner, M., Wirth, K., Konrad, A.,
Warneke, K. (2025).

1RM measurement in the
standing calf raise: Evaluation of reli-
ability and measurement error.

Journal of Applied Sports Sciences,
9(1), pp. 40 - 48.

DOI: 10.37393/JASS.2025.09.01.4



This work is licensed under
a Attribution-Non
Commercial-ShareAlike 4.0
International (CC BY-NC-SA 4.0)

ABSTRACT

Investigations into the maximum strength capacity of the plantarflexors have gained interest. An excellent intraclass correlation coefficient (ICC) for the standing calf raise-1RM has been reported. However, research has shown that ICC alone does not sufficiently reflect measurement reliability.

The aim of this study was to reassess the reliability of the standing calf raise-1RM. Fifty-six physical education students performed standing calf raises on two occasions. ICC, standard error of measurement (SEM), and minimal detectable change (MDC) were calculated, as well as mean absolute error, mean absolute percentage error, and typical error. A dependent t-test was performed to assess systematic error, and a Bland-Altman plot was created.

Low levels of both absolute and relative error were observed. Comparing the different measures, previous criticisms of the ICC and measures calculated from it, such as SEM and MDC, are highlighted. Researchers are advised to report various independent measures to illustrate the accuracy of their procedures.

Keywords: plantarflexion, maximum strength, accuracy, precision

INTRODUCTION

Maximal strength is considered crucial for high performance in various sporting tasks (Stone et al., 2024; Suchomel et al., 2016). Regarding the lower extremity, most studies and performance tests focus on the thigh musculature, specifically the quadriceps, hamstrings, and glutes (Chelly et al., 2010; McBride et al., 2009; Möck et al., 2018b; Möck et al., 2022; Suchomel et al., 2016). Although their role in achieving top-athletic performance is undisputed, the plantar flexors play an important role in proprioception, gait and sprinting, change of direction and jumping (Arch et al., 2018; Hubbuch et al., 2015; Jonkers et al.,

2003; Kadlubowski et al., 2024; Keiner et al., 2021; Möck et al., 2018a; Möck et al., 2023; Warneke et al., 2022b). However, only a few studies address issues that need to be considered when evaluating this comparatively small muscle group. The results of Möck et al. (2023) and Warneke et al. (2022b) suggest that the knee angle, and therefore the involvement of the m. gastrocnemius, is of high relevance for the relationship between maximum strength in the calf raise and jumping performance. Since the standing calf raise seems to produce the highest correlation coefficients with dynamic performance (Möck et al., 2023; Warneke et al., 2022b) and different maximum

strength tests for this muscle group cannot be used interchangeably (Warneke et al., 2022a), 1RM testing in the standing calf raise might be proposed as preferential method.

Despite the crucial role of internal and external validity in strength measurements (Warneke et al., 2023), ensuring testing reliability —i.e., repeatability and reproducibility— must be provided to assess the practical relevance of a measurement. Most commonly, consistency or agreement is quantified by the intraclass correlation coefficient (ICC) (Koo & Li, 2016). For the 1RM testing in the standing calf raise, the ICC has been stated by Möck et al. (2018a) and Möck et al. (2023) as $ICC = .987$. To supplement this measure, some authors propose the calculation of the standard error of measurement (SEM) or the minimal detectable change (MDC) to define the least amount of training induced change that needs to be reached to ensure that the increase is not a product of measurement errors (Impellizzeri et al., 2008; Martins et al., 2017; Wollin et al., 2016).

Despite common practice, such indices often neglect the role of factors such as learning effects (systematic bias) and random error, or noise (Hopkins, 2000). Since testing maximal strength is invalidated by learning effects (it would test how fast an individual learns, rather than determining the actual strength capacity), Hopkins (2000) and Atkinson & Nevill (2000) discussed additional statistical calculations to account for such sources of measurement error. Therefore, authors were encouraged to implement Bland Altman analyses (Atkinson & Nevill, 2000) or the typical error (TE) (Hopkins, 2000), as the sole focus on the ICC does not provide sufficient information whether an acceptable and practically relevant level of reliability is reached (Lamb, 1998). However, Grgic et al. (2020) claimed (for the 1RM squat test) that such recommendations

remain widely unaddressed and only a few articles performed agreement analyses. To supplement the graphical inspection of the measurement error, the calculation of the mean absolute error (MAE) and the mean absolute percentage error (MAPE) has been proposed (Kim & Kim, 2016; Willmott & Matsuura, 2005, 2006).

To close this gap in literature, the aim of this study was to evaluate the reliability of 1RM measurements in the standing calf raise, calculating multiple statistical measures for the measurement device to account for diverse error sources suggested by the literature.

METHODOLOGY

Participants

Fifty-six physical education students (40 males, 16 females) participated in this study. The mean age was 23.7 ± 3.0 years, the body height 176.9 ± 8.1 cm and the body weight 74.2 ± 10.3 kg (male: 23.95 ± 3.16 years, 180.18 ± 6.36 cm; 78.65 ± 8.32 kg; female: 23.00 ± 2.39 years, 168.75 ± 5.95 cm, 63.21 ± 5.39 kg). All the participants were recreationally to ambitiously active, participating in structured university sports courses and additionally training in their respective sports at least twice a week. Each participant was informed of the experimental risks involved with the research. All subjects provided informed written consent. The research design was approved by the institutional review board (German University of Health & Sport, No. 01/2019.92002800) and was conducted in accordance with the use of human subjects and the Declaration of Helsinki.

Experimental protocol

The 1RM in the standing calf raise was determined in two sessions separated by one week. The testing device of Möck et al. (2023) was used (Figure 1). The participants

performed an individual warm-up consisting of mobilization exercises and submaximal calf raises. The 1RM measurements were performed with the barbell placed on the shoulders above the 7th cervical vertebrae. The amplitude of the movement required a minimum distance of 4 cm from the neutral position in both directions of movement to be counted as a valid trial. The participants lowered the barbell 4 cm into dorsiflexion of the ankles, followed by 4 cm of plantarflexion concerning the neutral position. Therefore, the total

distance traveled by the barbell was 12 cm (4 cm eccentrically, 8 cm concentrically). The movement amplitude was measured by two laser distance systems with a measuring range of 2000 mm and a measuring rate of 750 Hz (Model: optoNCDT 1302-200, Manufacturer: Micro-Epsilon, Germany). An attempt was declared as invalid when the necessary range of motion in at least one direction of movement could no longer be generated, or the load was moved by using the knee and hip joints.

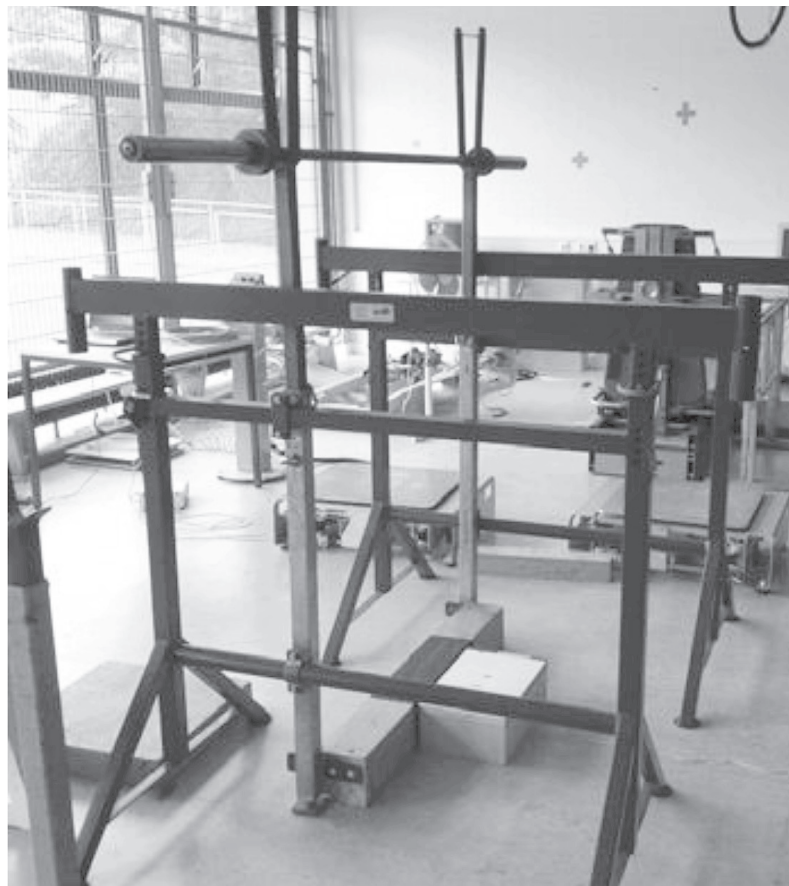


Figure 1. Apparatus for 1RM testing in the standing calf raise.

Statistical analysis

Data analysis was performed with SPSS 11.5 (SPSS, Inc., Chicago, IL, USA) and Microsoft Excel 365 (Microsoft Corporation, Redmond, WA, USA). Descriptive data is reported as mean (M) and standard deviation (SD) as well as maximum (Max) and mini-

mum (Min). To assess interday measurement reliability, the ICC was calculated from the 1RM values of the two testing sessions.

ICC thresholds were adopted and classified as follows: $< .50$ = poor, $.50 - .75$ = moderate, $.75 - .90$ = good, and $> .90$ = excellent (Koo & Li, 2016). As the SEM (Tighe et al.,

2010) and the MDC (Seamon et al., 2022) are well, we calculated those using the following considered standard reliability parameters as formulas:

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

and

$$MDC = SEM * 1.96 * \sqrt{2}$$

The TE, according to Hopkins, was calculated with the formula:

$$TE = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Additionally, a Bland-Altman agreement analysis was plotted, and the mean difference (systematic error) was calculated using the dependent two sample *t*-test (Atkinson & Nevill, 2000) with the respective LoAs provided to account for random scattering (Bland & Altman,

1986; Bland & Altman, 1999; Doğan, 2018). Additionally, the mean absolute error (MAE) (Willmott & Matsuura, 2005; Willmott & Matsuura, 2006) as well as the mean absolute percentage error (MAPE) (Kim & Kim, 2016) were calculated using the following formulas:

$$MAE = \frac{1}{n} * \sum_{i=1}^n |x_i - y_i|$$

and

$$MAPE = \frac{1}{n} * \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| * 100$$

RESULTS

The descriptive data of the two testing sessions are presented in Table 1. The participants lifted a mean 1RM in the standing calf raise of

191.92 ± 45.32 kg in the first testing session and 193.95 ± 45.03 kg in the second testing session.

Table 1. Descriptive data of the two testing sessions.

	Measurement 1	Measurement 2
Mean [kg]	191.92	193.95
SD [kg]	45.32	45.03
Max [kg]	306.20	306.20
Min [kg]	83.60	88.80

The interday reliability analysis revealed an ICC of 0.981, a SEM of 0.727 kg, and an MDC of 2.014 kg for the 1RM measurement in

the standing calf raise (Figure 2). Additionally, an MAE of 4.079 kg and an MAPE of 2.210 % could be observed.

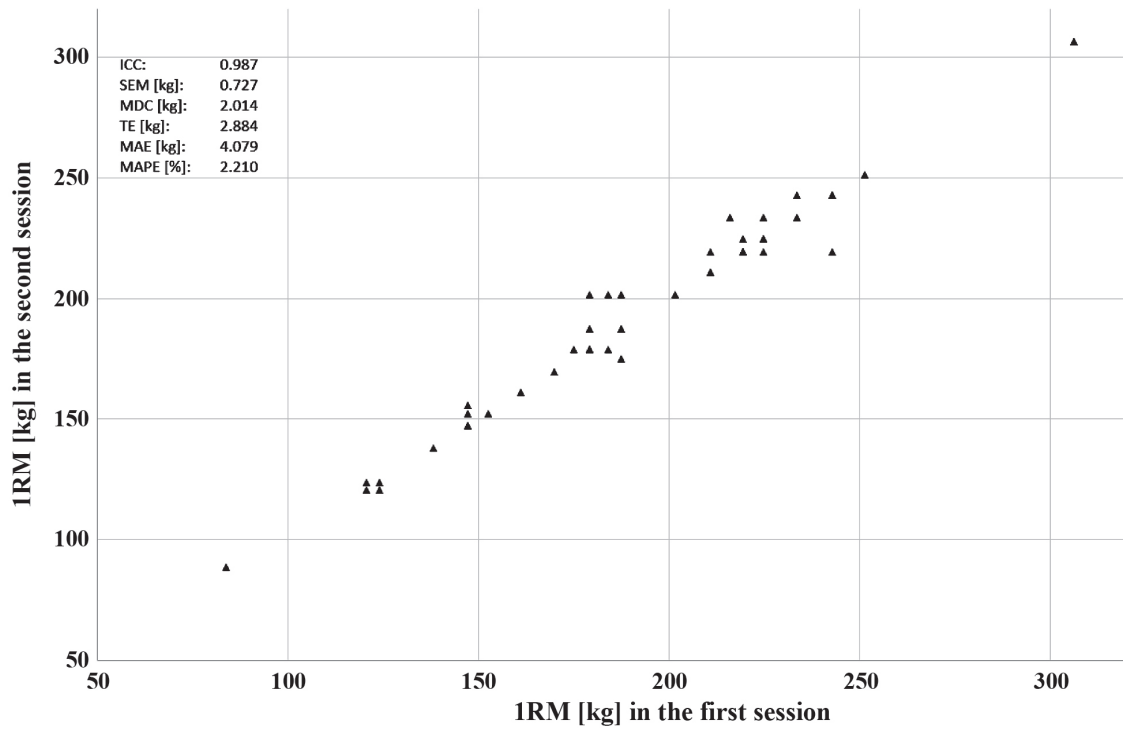


Figure 2. Scatterplot and reliability measures of the standing calf raise.

The Bland-Altman plot shows a difference indicating no systematic bias. The addition of the mean of 2.0 kg with the LoAs identified at -12.09 kg and 16.15 kg, respectively yielded no significant result. (Figure 3). The plot reveals a horizontal trend,

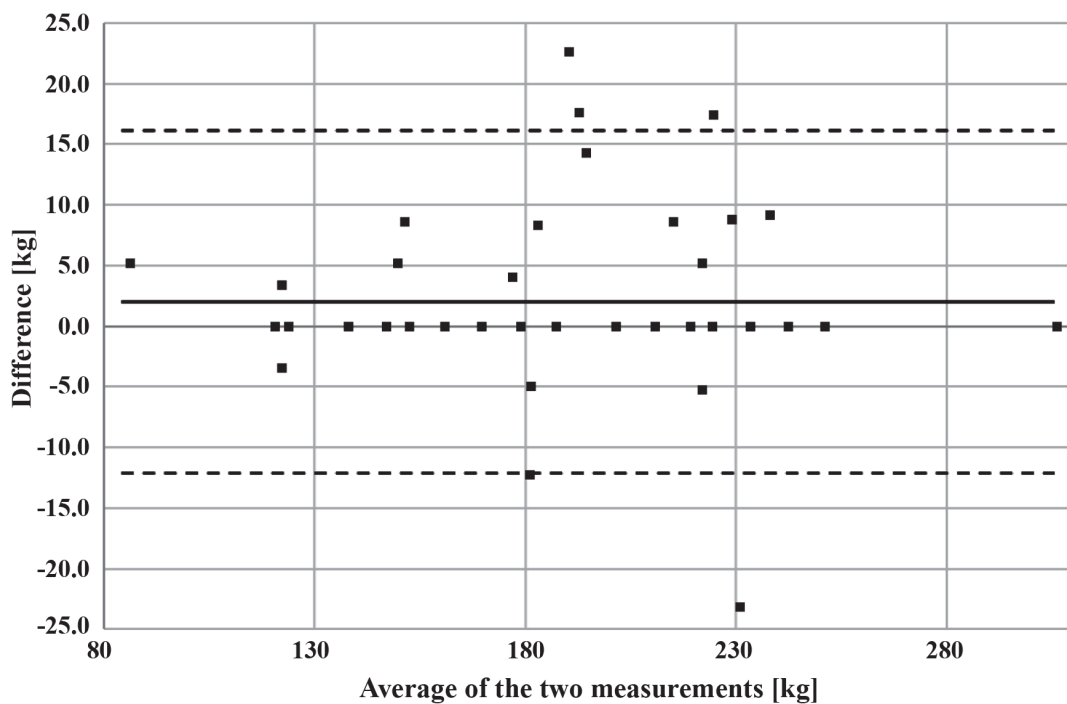


Figure 3. Bland-Altman plot of the standing calf raise measurements.

DISCUSSION

This study investigated the reliability of a previously reported method to measure the 1RM in the standing calf raise. An excellent ICC of .987 with a SEM of 0.727 kg was observed, accompanied by a MAE of 4.079 kg and a MAPE of 2.210%. Additionally, an MDC of 2.014 kg and a TE of 2.884 kg were obtained. The results demonstrate that the 1RM measurement in the standing barbell calf raise is a reliable method for investigating the maximum strength of the plantarflexors, as established guidelines (Koo & Li, 2016) suggest. Practitioners and researchers interested in investigating aspects of the dynamic maximum strength of the plantarflexors are recommended to utilize this procedure for research purposes and training monitoring. Nevertheless, 5 participants showed differences of more than 15 kg between the two measurements. This might be interpreted as insufficient familiarization in the case of greater 1RM values in the second testing. Therefore, researchers and practitioners should be aware that one familiarization session prior to testing might not be sufficient for all individuals (Dias et al., 2005).

This study's results provide interesting insights into the usefulness of different reliability measures. The SEM and the MDC are often advocated as indices of measurement reliability (Seamon et al., 2022; Tighe et al., 2010). However, the present results, with a mean error of MAE = 4.079 kg, contrasting with SEM = 0.727 kg and MDC = 2.014 kg, question the usefulness of the SEM and the MDC. If the mean error of a measurement is roughly double the minimal detectable change, the MDC cannot hold up to its name. This observation illustrates a mathematical problem with the SEM and the MDC. Both these measures are derived from the ICC of a measurement. Therefore, a high ICC value inevitably leads to low SEM and MDC values. As previous investigations

have shown, the ICC itself should be considered a nonsufficient and problematic measure of reliability, since an excellent ICC may be accompanied by considerable mean errors and mean percentage errors (Warneke et al., 2022c; Warneke et al., 2023). Additionally, the investigated sample exhibited a considerable level of heterogeneity in terms of the 1RM in the standing calf raise. As this can lead to an overestimation of correlation coefficients (Wilcox, 2013), the results further underline that the validity of the ICC as a measure of reliability must be questioned.

The results indicate that reporting single measures to indicate measurement reliability is not justified. Furthermore, researchers and practitioners need to be aware of the strengths and weaknesses of a broad spectrum of proposed reliability measures. Study or training results, including cross-sectional or longitudinal group comparisons, cannot be validly interpreted without knowledge of the measurement error and reliability, even if they reach statistical significance. Regarding the present results, supposedly significant differences of less than 4.079 kg or 2.210 % in plantarflexion 1RM should be treated as no difference because they do not exceed the mean error of this measurement method and therefore the area of uncertainty.

Regarding the estimation of effect sizes (ES), calculations based on the ICC have been proposed to correct study ES for the uncertainty of a measurement (Ahn et al., 2012). As pointed out above, these correction methods also need to be questioned regarding their problematic validity as an indicator of measurement reliability.

CONCLUSIONS

In conclusion, the measurement of the 1RM in the standing calf raise displays low levels of both absolute and percentage mean errors.

However, the results suggest that correlational measures of reliability should be viewed as problematic and supplemented with additional independent procedures to gain a comprehensive understanding of data quality and accuracy. Researchers are advised to report reliability data from comparable studies but also reproduce these analyses with their own sample to control for possible bias based on sample characteristics and investigator selection.

REFERENCES

- Ahn, S., Myers, N. D., & Jin, Y. (2012). Use of the estimated intraclass correlation for correcting differences in effect size by level. *Behavior Research Methods, 44*, 490–502.
- Arch, E. S., Colon, S., Silbernagel, K. G., & Crenshaw, J. R. (2018). Evaluating the relationship between gait and clinical measures of plantar flexor function. *Journal of Electromyography and Kinesiology, 43*, 41–47.
- Atkinson, G., & Nevill, A. (2000). Measures of Reliability in Sports Medicine and Science. *Sports Medicine, 30*(5), 375–381.
- Bland, J. M., & Altman, D. G. (1986). Statistical Methods of Assessing Agreement between Two Methods of Clinical Measurement. *Lancet, i*, 307–310.
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135–160.
- Chelly, M. S., Chérif, N., Amar, M. B., Hermassi, S., Fathloun, M., Bouhlel, E., ... & Shephard, R. J. (2010). Relationships of peak leg power, 1 maximal repetition half back squat, and leg muscle volume to 5-m sprint performance of junior soccer players. *Journal of Strength & Conditioning Research, 24*(1), 266–271.
- Dias, R. M. R., Cyrino, E. S., Salvador, E. P., Caldeira, L. F. S., Nakamura, F. Y., Papst, R. R., ... & Gurjão, A. L. D. (2005). Influence of familiarization process on muscular strength assessment in 1-RM tests. *Revista Brasileira de Medicina do Esporte, 11*, 34–38.
- Doğan, N. Ö. (2018). Bland-Altman Analysis: A Paradigm to Understand Correlation and Agreement. <https://doi.org/10.1016/j.tjem.2018.09.001>
- Grgic, J., Lazinica, B., Schoenfeld, B. J., & Pedisic, Z. (2020). Test–retest reliability of the one-repetition maximum (1RM) strength assessment: a systematic review. *Sports Medicine-Open, 6*, 1–16.
- Hubbuck, J. E., Bennett, B. W., & Dean, J. C. (2015). Proprioceptive feedback contributes to the adaptation toward an economical gait pattern. *Journal of Biomechanics, 48*(11), 2925–2931.
- Impellizzeri, F. M., Bizzini, M., Rampinini, E., Cereda, F., & Maffiuletti, N. A. (2008). Reliability of isokinetic strength imbalance ratios measured using the Cybex NORM dynamometer. *Clinical Physiology and Functional Imaging, 28*(2), 113–119.
- Jonkers, I., Stewart, C., & Spaepen, A. (2003). The complementary role of the plantarflexors, hamstrings, and gluteus maximus in the control of stance limb stability during gait. *Gait & Posture, 17*(3), 264–272.
- Kadlubowski, B., Keiner, M., Wirth, K., & Csapo, R. (2024). Association between Sprint and Jump Performance and Maximum Strength in Standing Calf Raise or Squat in Elite Youth Soccer Players. *Sports, 12*(4), 87.
- Keiner, M., Kadlubowski, B., Hartmann, H., Stefer, T., & Wirth, K. (2021). The influence of maximum strength performance in squats and standing calf raises on squat jumps, drop jumps, and linear as well as change of direction sprint performance in youth soccer players. *International Journal of Sports and Exercise Medicine, 7*(3), 190.
- Kim, S., & Kim, H. (2016). A New Metric of Absolute Percentage Error for Intermittent

- Demand Forecasts. *International Journal of Forecasting*, 32(3), 669–679.
- Koo, T. K., & Li, M. Y. (2016). A Guide-line of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
- Lamb, K. (1998). Test-Retest Reliability in Quantitative Physical Education Research: A Commentary. *European Physical Education Review*, 4(2), 145–152.
- Martins, J., Da Silva, J. R., Da Silva, M. R. B., & Bevilacqua-Grossi, D. (2017). Reliability and validity of the belt-stabilized handheld dynamometer in hip-and knee-strength tests. *Journal of Athletic Training*, 52(9), 809–819.
- McBride, J. M., Blow, D., Kirby, T. J., Haines, T. L., Dayne, A. M., & Triplett, N. T. (2009). Relationship between maximal squat strength and five, ten, and forty yard sprint times. *Journal of Strength & Conditioning Research*, 23(6), 1633–1636.
- Möck, S., Erlewein, S., Mickel, C., & Wirth, K. (2022). Relationship of Isokinetic Leg Press Force with jumping performance and 1RM in the squat. *Journal of Science in Sport and Exercise*, 4, 180–187.
- Möck, S., Hartmann, R., Wirth, K., Rosenkranz, G., & Mickel, C. (2018a). Correlation of dynamic strength in the standing calf raise with sprinting performance in consecutive sections up to 30 meters. *Research in Sports Medicine*, 26(4), 474–481.
- Möck, S., Hartmann, R., Rosenkranz, G., Mickel, C., & Wirth, K. (2018b). Maximal strength in the deep back squat correlates with sprinting performance over short distances. *International Journal of Applied Sports Sciences*, 30(2), 200–207.
- Möck, S., Hartmann, R., & Wirth, K. (2023). Vertical jumping performance relates to the one-repetition maximum in the standing calf raise and in the squat. *German Journal of Exercise and Sport Research*, 53(2), 139–147.
- Seamon, B. A., Kautz, S. A., Bowden, M. G., & Velozo, C. A. (2022). Revisiting the Concept of Minimal Detectable Change for Patient-Reported Outcome Measures. *Physical Therapy*, 102(8). <https://doi.org/10.1093/ptj/pzac068>
- Stone, M. H., Hornsby, G., Mizuguchi, S., Sato, K., Gahreman, D., Duca, M., ... & Haff, G. G. (2024). The Use of Free Weight Squats in Sports: A Narrative Review—Squatting Movements, Adaptation, and Sports Performance: Physiological. *Journal of Strength & Conditioning Research*, 38(8), 1494–1508.
- Suchomel, T. J., Nimphius, S., & Stone, M. H. (2016). The importance of muscular strength in athletic performance. *Sports Medicine*, 46, 1419–1449.
- Tighe, J., McManus, I., Dewhurst, N. G., Chis, L., & Mucklow, J. (2010). The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP(UK) examinations. *BMC Medical Education*, 10(1), 40.
- Warneke, K., Hillebrecht, M., Wirth, K., Schiemann, S., & Keiner, M. (2022a). Correlation between Isometric Maximum Strength and One Repetition Maximum in the Calf Muscle in Extended and Bended Knee Joint. *International Journal of Applied Sports Sciences*, 34(1), 61–71.
- Warneke, K., Keiner, M., Lohmann, L. H., Hillebrecht, M., Wirth, K., & Schiemann, S. (2022b). The Influence of Maximum Strength Performance in Seated Calf Raises on Counter Movement Jump and Squat Jump in Elite Junior Basketball Players. *Sport Mont*, 20(2), 63–68.
- Warneke, K., Keiner, M., Lohmann, L. H., Brinkmann, A., Hein, A., Schiemann, S., & Wirth, K. (2022c). Critical evaluation of commonly used methods to determine the concor-

dance between sonography and magnetic resonance imaging: a comparative study. *Frontiers in Imaging, 1*, 1039721.

Warneke, K., Wagner, C. M., Keiner, M., Hillebrecht, M., Schiemann, S., Behm, D. G., ... & Wirth, K. (2023). Maximal strength measurement: A critical evaluation of common methods—a narrative review. *Frontiers in Sports and Active Living, 5*, 1105201.

Wilcox, R. R. (2013). *Introduction to robust estimation and hypothesis testing*. Amsterdam: Elsevier.

Willmott, C. J., & Matsuura, K. (2005). Ad-

vantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Climate Research, 30*, 79–82.

Willmott, C. J., & Matsuura, K. (2006). On the Use of Dimensioned Measures of Error to Evaluate the Performance of Spatial Interpolators. *Int J Geogr Inf Sci, 20*(1), 89–102.

Wollin, M., Purdam, C., & Drew, M. K. (2016). Reliability of externally fixed dynamometry hamstring strength testing in elite youth football players. *Journal of Science and Medicine in Sport, 19*(1), 93–96.

Corresponding author:

Sebastian Puschkasch-Möck

Olympic Training and Testing Center of Hessen,
Frankfurt, Germany
Email: smoeck@lsbh.de