Study design: test-retest reliability of different methods of measuring the endurance of advanced to elite climbers

This paper describes the design of a study aimed at evaluating the test-retest reliability of different methods of measuring forearm endurance of sports climbers. Subjects will be recruited and tested on two occasions, and data will be analyzed to evaluate intra-class correlation coefficients and coefficients of variation. This foundational work will help future data analysis and study design by letting researchers evaluate their study power, type I and II error rates, as well as effect size in a more rigorous manner.

Introduction

The growing popularity of sports climbing, culminating in its addition as an olympic discipline, has led to a rapid increase in research interest around training plans and predictors of climbing performance.

Several studies have been conducted to investigate the effect of different protocols on the strength and endurance of climbers. The training methods that have been examined to assess the resulting change in strength, rate of force development or endurance of the finger flexors include submitting climbers to training with varying grip types [Levernier2019], edge sizes [Lopez-Rivera2012, Hermans2022], or load [Hermans2016, Lopez-Rivera2019]. Some studies have compared the effect of distinct protocols on both strength and endurance [Devise2022].

Researchers have used a complex array of measurement methods to evaluate these traits. For example, strength is sometimes measured as the peak force applied by climbers on a force sensor during a contraction [Rokowski2021], while others use the maximum load that a climber can lift with a certain grip type for an arbitrary amount of time, or the average over such period [Giles2019, Devise2022]. What’s more, the measured strength seems to depend on the specific angle of elbow flexion and shoulder engagement that test subjects are instructed to use [Michailov2018].

Measuring endurance comes with its own set of issues because of the distinction between anaerobic, and aerobic metabolic regimes, and one has to consider the validity of the chosen construct in discriminating between them. Nonetheless, different methods have been employed in an attempt to assess these forms of endurance before and after training, or simply to correlate them to climbing performance. Maciejczyk et al. [Maciejczyk2022] measured the maximum hang time of climbers on 23mm rung at 60% +/- 10% MVC. Ozimek et al. [Ozimek2017] measured the maximum hang time with two hands on 25 and 40mm rungs. Some also measured the number of consecutive hangs at bodyweight with specified contraction-relaxation time [Michailov2018]. Drawing from literature in endurance sports, other researchers have proposed the use of either intermittent or continuous all-out tests. One such all-out test which has gained popularity in recent years is the 4 min all-out test employed by Giles et al. [Giles2019]. More recently, Devise et al. [Devise2022] tested athletes on a series of

10s contractions with 6s of resting in-between at 80% maximum force until they cannot maintain it, at which point they are instructed to exert the maximal effort, for a total of 24 contractions.

This shows that despite a large body of work, there is still a lack of standard methods in measuring various aspects of performance in climbing. A crucial criterion for the choice of a method is its test-retest reliability. It has been shown that the reliability of a measurement (or its lack thereof) can largely influence statistical analyses: when doing multiple regression, errors of measurement can massively inflate the false positive [Shear2013, Westfall2016] and false negative rates (statistical power) [Kayongo2007, Zimmerman2015], beyond what one would intuitively expect from the p-values. Analysis of variance (ANOVA), a commonly used tool in sports science which is a form of regression, is not shielded from these issues. Measurement error also impacts the estimation of the effect size, potentially leading to either underestimate or overestimate [Loken2017].

Reliability is also important for coaches: poor reliability leads to larger errors of estimate of the true value of an athlete’s test score, which can make it hard to assess the efficacy of a given training routine.

Several authors have estimated the test-retest reliability of different methods of strength measurement and have typically found excellent reliability [Torr2022] [Michailov2018][Levernier2019]. Much less work has been done on endurance measurements. Michailov et al. have assessed the reliability of the “fatigue index” after a 30s all out test, as well as that of the number of repetitions in a target zone of 60% +/- 10% of the maximum voluntary contraction (MVC) with 8 seconds - 2 seconds contraction - relaxation time, and found relatively reliability [Michailov2018]. McClean et al. evaluated the test-retest reliability of the 4 minutes all-out test developed by Giles et al [McClean2023].

These groups found reliabilities typically considered “good”, but when considered in light of the impact of reliability on false discovery, false positives, and false negative rates, this characterization may not be perfectly accurate. We also found that to this day, there has been no evaluation in literature of the test-retest reliability of the continuous hang time at a certain % MVC, or of the measurement method used by Devise et al.

The aim of this study is to evaluate the reliability of different methods of assessing climbers endurance: continuous hang time at 60% MVC, 80% MVC, as well as the Devise et al. protocol. There will be no discussion of the validity of these constructs, or their correlation to sports climbing performance, but all data will be accessible freely for further use.

Methods

General procedures for the tests

After determining the sample size for acceptable statistical power, advanced to elite climbers (at least 3 repetitions of outdoor routes graded 19 to 25 on the IRCRA reporting scale, or French 7b to 8b) will be recruited. Subjects will be required to have experience fingerboarding (at least 3 times per month the past 3 months). Relevant information on the climbers will be reported following guidelines set out by the IRCRA [Draper2015], with the exception of the maximum onsight grade which will be collected using the past 6 months instead of 3. Data acquisition is planned to be made using the Climbro force sensor. Subjects will be randomly assigned to one of the three groups, and undergo a familiarization session with the device used as well as the max strength test and the protocol that they will be tested on. During the familiarization session, climbers will be asked to choose one grip type to use for all future tests: half-crimp (4 fingers contact, 90° flexion of the PIP joint with 0° flexion at the DIP and MCP joints) or open crimp (4 fingers contact, 0° flexion at the MCP joint, and relaxed constraint at the PIP and DIP joint, no DIP extension). Subjects will hang on the Climbro rung with elbows slightly bent so that shoulders are engaged, thereby reducing the risk of injury as proposed by Giles et al. [Giles2019]. If their maximum force exceeds their bodyweight, they will be weighted using a weight-belt and sufficient added weights such that they cannot lift off the ground. All sessions will be preceded by a standardized warm-up protocol consisting of both nonspecific cardio exercises and specific exercises, the details of which will be reported in the final paper. Two measurements will be realized with a 7 days interval and similar instructions given for the 24 hours preceding the respective tests. The exact instructions are yet to be determined and will be reported in the final paper. Apart from these instructions, subjects will be asked to maintain their usual climbing routine between the two tests. The reliability of the test scores will be expressed as an intra-class correlation coefficient on repeated single measurements with the same raters (ICC(3,1)) [Weir2005] . All hangs will be realized on the 20mm rung in an attempt to minimize the influence of slight misplacement by subjects or changes in conditions influencing friction. More details on the different protocols and instructions given to climbers when executing them is given below.

Maximum voluntary contraction test

Subjects hang with both hands and slightly bent elbows on the 20mm rung and apply as much force as possible for 6 seconds. The average force over the duration of the contraction is considered as their MVC. Note that Devise et al. used the Smartboard which has one force sensor per hand compared to a single sensor for the Climbro system planned to be used for this study.

60% and 80% MVC continuous test

Subjects are tasked to maintain a target intensity of 60% (80%) of their MVC with a 10% tolerance for as long as possible. The total time under tension in the target zone is considered their continuous endurance at 60 (80%) MVC, or C60e (C80e) for short.

Devise et al. test

Subjects perform a total of 24 contractions of 10 seconds, with 6 seconds of rest in between. At first, they are asked to maintain an intensity of 80% with a 10% tolerance. Once they fall outside of the tolerance band, they are asked to apply as much force as possible for the remaining contractions. This allows us to extract two quantities named *stamina* and *endurance* by Devise et al. The percentage of stamina is the percentage of contractions that were in the target zone. The endurance is the average of the 5 lowest intensity contractions, after excluding the poorest contraction.

Reliability calculations

Reliability calculations will be conducted following the procedures described by Weir [Weir2005]. Specialized software may be used and the help of a statistician may be sought to validate our results.

Expected results and discussions

The reliability of these test scores is expected to be similar to that of already assessed endurance tests, namely in the 0.75-0.85 range for the ICC. Using these results, commentary of the statistical analysis of Devise et al. will be provided as an example of the utility of using measurement methods of known test-retest reliability.

Future work

Knowing the test-retest reliability of more endurance tests will allow more research to be conducted in order to find optimal training regimens to improve performance in this domain. Much work is also left to do on the validity of these tests and their relationship with both climbing performance, and different metabolic regimes.

References

[Levernier2019] Levernier, G., & Laffaye, G. (2019). Four weeks of finger grip training increases the rate of force development and the maximal force in elite and top world-ranking climbers. *The Journal of Strength & Conditioning Research*, *33*(9), 2471-2480.

[LopezRivera2012] López-Rivera, E., & González-Badillo, J. J. (2012). The effects of two maximum grip strength training methods using the same effort duration and different edge depth on grip endurance in elite climbers. *Sports Technology*, *5*(3-4), 100-110.

[Hermans2022] Hermans, E., Saeterbakken, A. H., Vereide, V., Nord, I. S., Stien, N., & Andersen, V. (2022). The effects of 10 weeks hangboard training on climbing specific maximal strength, explosive strength, and finger endurance. *Frontiers in Sports and Active Living*, *4*, 888158.

[Hermans2016] Hermans, E., Andersen, V., & Saeterbakken, A. H. (2017). The effects of high resistance–few repetitions and low resistance–high repetitions resistance training on climbing performance. *European journal of sport science*, *17*(4), 378-385.

[LopezRivera2019] López-Rivera, E., & González-Badillo, J. J. (2019). Comparison of the effects of three hangboard strength and endurance training programs on grip endurance in sport climbers. *Journal of human kinetics*, *66*(1), 183-195.

[Devise2022] Devise, M., Lechaptois, C., Berton, E., & Vigouroux, L. (2022). Effects of different hangboard training intensities on finger grip strength, stamina, and endurance. *Frontiers in Sports and Active Living*, *4*, 862782.

[Rokowski2021] Rokowski, R., Michailov, M., Maciejczyk, M., Więcek, M., Szymura, J., Draga, P., ... & Szygula, Z. (2024). Muscle strength and endurance in high-level rock climbers. *Sports biomechanics*, *23*(8), 1057-1072.

[Giles2019] Giles, D., Hartley, C., Maslen, H., Hadley, J., Taylor, N., Torr, O., ... & Fryer, S. (2021). An all-out test to determine finger flexor critical force in rock climbers. *International Journal of Sports Physiology and Performance*, *16*(7), 942-949.

[Michailov2018] Michailov, M. L., Baláš, J., Tanev, S. K., Andonov, H. S., Kodejška, J., & Brown, L. (2018). Reliability and validity of finger strength and endurance measurements in rock climbing. *Research quarterly for exercise and sport*, *89*(2), 246-254.

[Maciejczyk2022] Maciejczyk, M., Michailov, M. L., Wiecek, M., Szymura, J., Rokowski, R., Szygula, Z., & Beneke, R. (2022). Climbing-specific exercise tests: energy system contributions and relationships with sport performance. *Frontiers in Physiology*, *12*, 787902.

[Ozimek2017] Ozimek, M., Rokowski, R., Draga, P., Ljakh, V., Ambroży, T., Krawczyk, M., ... & Mucha, D. (2017). The role of physique, strength and endurance in the achievements of elite climbers. *PLoS One*, *12*(8), e0182026.

[Shear2013] Shear, B. R., & Zumbo, B. D. (2013). False positives in multiple regression: Unanticipated consequences of measurement error in the predictor variables. *Educational and Psychological Measurement*, *73*(5), 733-756.

[Westfall2016] Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PloS one*, *11*(3), e0152719.

[Kayongo2007] Kanyongo, G. Y., Brook, G. P., Kyei-Blankson, L., & Gocmen, G. (2007). Reliability and statistical power: How measurement fallibility affects power and required sample sizes for several parametric and nonparametric statistics. *Journal of Modern Applied Statistical Methods*, *6*, 81-90.

[Zimmeman2015] Zimmerman, D. W., & Zumbo, B. D. (2015). Resolving the issue of how reliability is related to statistical power: adhering to mathematical definitions. *Journal of Modern Applied Statistical Methods*, *14*, 9-26.

[Loken2017] Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, *355*(6325), 584-585.

[Torr2022] Torr, O., Randall, T., Knowles, R., Giles, D., & Atkins, S. (2022). Reliability and validity of a method for the assessment of sport rock climbers' isometric finger strength. *The Journal of Strength & Conditioning Research*, *36*(8), 2277-2282.

[McClean2023] McClean, Z. J., MacDougall, K. B., Fletcher, J. R., Aboodarda, S. J., & Macintosh, B. R. (2023). Test-Retest Reliability of a 4-Minute All-Out Critical Force Test in Rock Climbers. *International Journal of Exercise Science*, *16*(4), 912.

[Weir2005] Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research*, *19*(1), 231-240.

[Draper2015] Draper, N., Giles, D., Schöffl, V., Konstantin Fuss, F., Watts, P., Wolf, P., ... & Abreu, E. (2015). *Comparative grading scales, statistical analyses, climber descriptors and ability grouping: International Rock Climbing Research Association position statement. Sports Technology, 8 (3–4), 88–94*.