Intra and Inter Day Reliability of Six-Second WattBike Performance Metrics in Elite Rugby Union Players

Adam Grainger [1], **Ross D. Neville** [1]

[1] Institute of Sport and Health, University College Dublin, Ireland

Rugby Union | Neuromuscular Function Testing | Reliability

Headline

The ability to generate force quickly is a key performance metric for athletes competing in rugby union match-play (1-3), with the assessment of force production in elite union rugby settings predominately involving jump testing (4-6). Cycle sprint testing provides a safe non-weight-bearing alternative; however, little is known about the reliability and practical utility of cycle sprint testing in this context.

Aim

Practitioners working in team field sports recognise that running-specific activity is challenging to implement in 'real world' settings (7,8). Moreover, in light of the reliability and feasibility issues reported with commonly-used jump testing1, researchers have begun to consider the utility of maximal intensity cycle testing as a means of monitoring fatigue (9-11). Whilst ecological validity is compromised to some extent with stationary assessments of force production, cycle testing has the advantage of being more reproducible than running tests performed outside, which are affected by weather and ground conditions (8). Additionally, it has been recognised that sprint cycle testing assesses a speed-power element of movement that is overlooked by jump testing (12).

Research in Australian Rules football has shown high reliability within a single session for six-second WattBike performance tests, with recommendations for its use to track meaningful changes in performance over time (9). However, to date, there is only one study (11) addressing the reliability and practical utility of Wattbike testing within elite-level rugby union. Moreover, the study reports data which leads to an unrealistically low smallest worthwhile change (10.68 and 7.5 W for six- and thirty-second tests of peak power, respectively) which, when considered alongside their reported typical error of measurement (54.8 and 67.8 W), results in a unacceptable signal-to-noise ratio (i.e. well in excess of the accepted threshold of $\leq$1) (7,13,14).

Clearly, there is scope for researchers to assess the reliability of the different possible metrics that can be extracted from Wattbike Pro performance testing with elite rugby union athletes. For practitioners to confidently interpret changes in Wattbike testing following training or match-play, they need to be able to determine whether such changes are real, or simply an artefact of measurement error. Therefore, the aim of this study was to assess the absolute and relative reliability of the six-second Wattbike test in professional rugby union players.

Design

We assessed four different six-second Wattbike metrics (total distance completed, average power generated, peak power achieved, and peak power relative to body mass) both within a single session on the same day (intra-day) and between two sessions across different days (inter-day) with a week in between. For the assessment of intra-day reliability, subjects were required to complete two six-second WattBike tests (separated by one minute) upon the same WattBike on the same testing day. For the assessment of inter-day reliability, subjects completed two six-second WattBike tests across two testing days with a week separating each test.

Method

Independent samples of subjects were used for intra-day (n=15) (age 27.5 ± 5.1, training age 8.5 ± 5.0, height 185.8± 7.2 cm, weight 100.1 ± 12.7 kg) and inter-day (n=24) (age 27.5 ± 5.1, training age 8.5 ± 5.0, height 185.8 ± 7.2 cm, weight 100.1 ± 12.7 kg) testing. Subjects were from the same professional rugby club, covered all playing positions, had no current injury issues, and adhered to the same training program.

Testing was performed on a single factory-calibrated WattBike Pro cycle ergometer (Wattbike Pro, Nottingham, United Kingdom). The seat and handles were set by the subjects based on individual preference, which remained consistent across all testing sessions. The placement of the six-second WattBike test within a subject's training day (between 8:00 and 10:00am) was standardised to minimise the effect of fatigue. Subjects performed ten minutes of dynamic stretching and five minutes steady-state cycling on another WattBike before testing. When completing the sprint cycle trial, subjects were instructed to start from a stationary position and to remain seated throughout. Subjects were also instructed to produce a maximal effort for the duration of the test.

The Wattbike was 'zeroed' prior to each testing day to ensure reliability in accordance with the manufacturer's recommendations. The WattBike Pro cycle ergometer's performance computer calculated power via a load cell located next to the chain. As force is exerted through the cranks, the load cell calculates power by the sum of all forces applied during one complete pedal revolution. The power output in Watts (W) that is presented on the WattBike monitor upon completion of the test is calculated using the following formula: P(W)=(F[N]*1[M])/T[S] Peak power (W), mean power (W), relative peak power (W/kg), and distance completed (m) throughout the six-second WattBike test were recorded. Participants were encouraged to achieve peak power as early as possible in the trial, as research (15) suggests that individuals who reach peak power quickly are also those with the highest peak power values.

Testing was conducted three days post-match to minimise the influence of game-induced fatigue. Given that the subjects were conditioned for elite-level competitive match-play, it is unlikely that game- and training-induced fatigue affected the results. No additional recovery strategies were administered during the testing period, and dietary intake was not manipulated for the purposes of this study.

Table 1. Pre- and post-test descriptive statistics for cycle sprint ergometer within- and between sessions, with magnitude-based inferences for the differences in means.

	Pre-test Mean ± SD	Post-test Mean ± SD	Effect[a] ×/÷ 95% CL	Inferences	p value
Within-session Reliability					
Total Distance Covered (m)	121.28 ± 5.92	120.98 ± 5.91	1.001 ×/÷ 1.004	Trivial****	0.66
Mean Power (W)	1250.26 ± 156.95	1254.67 ± 150.90	1.004 ×/÷ 1.017	Trivial****	0.63
Peak Power (W)	1462.85 ± 224.92	1477.60 ± 221.79	1.001 ×/÷ 1.014	Trivial****	0.11
Relative Power (W/kg)	14.52 ± 1.58	14.52 ± 1.63	1.00 ×/÷ 1.01	Trivial****	0.91
Between-session Reliability					
Total Distance Covered (m)	118.25 ± 6.32	117.40 ± 5.44	0.99 ×/÷ 1.03	Trivial****	0.52
Mean Power (W)	1277.33 ± 138.20	1279.25 ± 147.59	1.001 ×/÷ 1.04	Trivial****	0.93
Peak Power (W)	1492.88 ± 183.49	1471.17 ± 166.11	0.99 ×/÷ 1.02	Trivial****	0.09
Relative Power (W/kg)	14.31 ± 1.76	14.12 ± 1.62	0.99 ×/÷ 1.02	Trivial****	0.12

Abbreviations: SD, standard deviation; CL, confidence limits; m, metres; W, peak power in watts; kg, kilograms. [a]Paired samples t-tests were used to assess pre-post changes for each of the four cycle sprint metrics. Outcomes were evaluated as factor changes in the mean using the following thresholds: <1.11, trivial; 1.11-1.43, small; 1.43-2.0, moderate; 2.0-3.3, large; >3.3, very large. [b]Asterisks indicate effects clear at the 95% level and likelihood that the true effect is substantial, as follows: *possibly, **likely, ***very likely, ****most likely (17).

Table 2. Absolute and relative reliability outcomes for within- and between-session testing, with magnitude-based inferences.

	Within-session Reliability		Between-session Reliability	
	Estimate ± 95% CL	Inference	Estimate ± 95% CL	Inference
Total Distance Covered				
ICC[a] (r)	0.99 ± 0.01	Almost perfect	0.50 ± 0.31	Moderate
TEM[b] (raw)	0.57 ± 0.21	Trivial****	4.46 ± 1.79	Moderate****
CV[c] (%)	0.50 ± 0.34	Trivial****	3.79 ± 1.52	Trivial**
SWC[b] (raw)	1.18 ± 0.50		1.17 ± 0.37	
Signal-to-noise ratio[d]	0.48	Good	3.80	Poor
Mean Power				
ICC[a] (r)	0.98 ± 0.03	Almost perfect	0.75 ± 0.19	Very large
TEM[b] (raw)	24.66 ± 9.20	Trivial****	73.02 ± 29.26	Small****
CV[c] (%)	2.14 ± 0.80	Trivial****	5.79 ± 2.32	Small**
SWC[b] (raw)	44.67 ± 18.87		28.56 ± 8.93	
Signal-to-noise ratio[d]	0.55	Good	2.56	Poor
Peak Power				
ICC[a] (r)	0.99 ± 0.01	Almost perfect	0.95 ± 0.05	Almost perfect
TEM[b] (raw)	23.57 ± 8.79	Trivial****	42.24 ± 16.92	Trivial*
CV[c] (%)	1.77 ± 0.66	Trivial****	2.98 ± 1.1.19	Trivial****
SWC[b] (raw)	44.67 ± 18.87		34.92 ± 10.92	
Signal-to-noise ratio[d]	0.53	Good	1.21	Marginal
Relative Power				
ICC[a] (r)	0.99 ± 0.01	Almost perfect	0.94 ± 0.05	Almost perfect
TEM[b] (raw)	0.14 ± 0.05	Trivial****	0.42 ± 0.17	Small**
CV[c] (%)	0.99 ± 0.37	Trivial****	2.95 ± 1.18	Trivial****
SWC[b] (raw)	0.32 ± 0.14		0.34 ± 0.11	
Signal-to-noise ratio[d]	0.43	Good	1.24	Marginal

Abbreviations: CL, confidence limits; ICC, intraclass correlation coefficient; TEM, typical error of measurement; CV, coefficient of variation; SWC, smallest worthwhile change. [a]Magnitude thresholds for deriving inferences about the ICC: <0.1, trivial; 0.1-0.3, small; 0.3-0.5, moderate; 0.5-0.7, large; 0.7-0.9, very large; 0.9-1.0, almost perfect. [b]Magnitudes for inferences about the TEM are based on SWC. Thresholds are derived based on the logic of thresholds for standardized effects outlined in Hopkins et al: <SWC, trivial; SWC-3*SWC, small; 3*SWC-6*SWC, moderate; >6*SWC, large. [c]Magnitude thresholds for deriving inferences about the CV are based on a SWC of 5% and follow a similar logic: <5%, trivial; 5-15%, small; 15-30%, moderate, >30%, large (17). [d]Signal-to-noise ratio represents TEM as a factor of SWC, with a factor of ≤1 designating acceptable practical utility (7).

In line with formulas presented in widely-cited reliability resources (16), we estimated the typical error of measurement (TEM), intraclass correlation coefficient (ICC), coefficient of variation (CV), and smallest worthwhile change (SWC) for total distance covered (m), mean power (W), peak power (W), and relative mean power (W/kg). Practical utility was assessed by evaluating the TEM as a factor of the SWC, with a factor of ≥1 designating acceptable practical utility (17).

Results

Table 1 shows pre- and post-test descriptive statistics for cycle sprint ergometer within- and between sessions. Table 2. Absolute and relative reliability outcomes for within- and between-session testing, with magnitude-based inferences are presented in Table 2.

Discussion

The main finding of this study was the almost perfect relative reliability (ICC >0.98) and satisfactory absolute reliability (CV <5%) of each metrics, from an intra-day perspective (i.e., within a single session on the same day). The signal-to-noise ratio (<1) also confirms the practical utility of each of these metrics from an intra-day perspective; something suggested, but not fully addressed, in previous studies (9).

Not only do the results of this study compare favourably with other sprint cycle tests, they also corroborate reliability data presented in other studies using six-second WattBike testing in similar field sports. For example, in Australian Rules football, Wehbe et al. (9) reported excellent relative and absolute reliability for peak power output in both academy (ICC = 0.97; CV = 2.4%) and elite adult (ICC = 0.96; CV = 3%) players. They report no raw data for the SWC, which prohibits further exploration of the signal-to-noise ratio (and therefore practical utility) of the six-second WattBike test in this context. In their study of amateur rugby players, Mathieu et al. (18) present comprehensive reliability data, which indicates that the six-second WattBike test is, at best, only marginally practically useful in an amateur setting. In fact, their study provides an excellent example of the need to assess relative and absolute reliability alongside the signal-to-noise ratio when conducting reliability analyses. From an intra-day perspective, they report high reliability outcomes for both peak (ICC = 0.96; CV = 2.6%) and mean (ICC = 0.96; CV = 2.3%) power output. However, when considered alongside the SWC, the signal-to-noise ratio is right on acceptable threshold of 1 for both assessment outcomes. In the amateur rugby context, the reliability of the intra-day six-second Wattbike cycle test is therefore marginal at best. Assessment of their data from an inter-day perspective (peak power: ICC = 0.94; CV = 2.3%; signal-to-noise ratio = 1.3) (mean power: ICC = 0.94; CV = 3.4%; signal-to-noise ratio = 1.4) shows that the test is probably too noisy to be practically useful.

The most comparable study for discussion against our data is Ripley et al. (11). Despite incorrectly using the formula for the reliability limits of agreement to calculate their SWC, they provide the only other comprehensive data on the inter-day reliability and practical utility of the six-second Wattbike cycle test for testing peak power output in elite level (professional) rugby union players. Similar to the outcomes presented in this study, Ripley et al. (11) report acceptable relative (ICC = 0.86) and absolute (CV = 3.8%) reliability for assessments of peak power between session separate by a week in between. The also report data which enables the calculation of a signal-to-noise ratio that is above the acceptable threshold (>1). However, we would question the baseline standard deviation reported in that study, that we used to derive a SWC – a signal-to-noise ratio of 5.13 seems unlikely in the case of acceptable relative and absolute reliability. If nothing else our data provides a confirmatory evidence of their reliability data and a more realistic assessment of the practical utility of assessing peak power output between sessions using the six-second Wattbike test. The practical utility of inter-day Wattbike testing is marginal (as shown by our signal-to noise ratios in Table 2). It is surely not practically useless, which one would have to conclude based on a signal-to-noise ratio of 5.13. Ripley et al. (11) provide no intra-day reliability to allow further comparison.

One possible explanation for the difference in our intra and inter-day findings – specifically, the lack of inter-day reliability, and lack of practical utility for Wattbike testing across a period of seven days – are possible differences in task engagement across these two types of trials. Maximal cycling tests involve a large element of task engagement. Mental focus prior to performing a tests such as the six-second WattBike test is key (19, 20). Whilst our protocol addressed possible moderating factors, and we found no meaningful differences in the subjects' physical or self-reported mental 'readiness' across the study period, a deeper analysis of psychometric differences across time would be required to rule out whether the motivational climate between trials was an important mediating factor. Incorporation of validated motivational climate inventories could provide another psychosocial layer of analysis that, to date, has not been much of a factor in strength and conditioning research reliability studies. Despite having shown that both field and gym load remained stable for all subjects across the assessment period, given Roe et al.'s (21) findings about the lack of sensitivity of a six-second Wattbike test to detect neuromuscular fatigue following high training loads, monitoring the effect of accumulating fatigue during competitive periods using non-invasive/non-intensive markers (heart rate variability, resting cortisol concentrations) is also worth considering for future studies.

Practical Applications

- The results of this study (considered alongside the minimal familiarization time needed to accustom athletes to cycle sprint testing) indicate that the six-second WattBike Pro test can be considered a reliable and useful complementary means of assessing power output in elite rugby union.
- These results also suggest that the six-second WattBike Pro test could be used (at the very least) as an initial reference of power output. Given widely-reported issues with confounding in jump testing due to jump strategy change, strength and conditioning professionals can consider cycle-sprint testing as both a practically useful component of their short-term performance assessment and optimisation arsenal.
- Stick to intra-day testing with WattBike Pro; inter-day testing is too noisy to be practically useful.

Limitations

- Subjects self-selecting WattBike fan resistance makes comparison between studies difficult.
- With WattBike Pro, peak values can be affected by the speed of performance, since the angular velocity used to calculate power is measured twice per revolution. The algorithm associated with the calculation of both speed and distance requires further investigation to more accurately assess peak, mean and distance values.

Twitter: Follow Adam Grainger @atgrainger

References

1.Argus CK, Gill ND, Keogh JWL. Characterisation of the differences in strength and power between different levels of competition in rugby union athletes. J Strength Cond Res. 2012;26(10):2698-2704.

2.Haff GG, Stone MH. Methods of developing power with special reference to football players. Strength Cond J. 2015;37(6):2-16.

3.Jones TW, Smith A, Macnaughton LS, French DN. Strength and conditioning and concurrent training practices in elite rugby union. J Strength Cond Res. 2016;30(12):3354-3366.

4.Kennedy R, Drake D. The effect of acute fatigue on countermovement jump performance in rugby union players during preseason training. J Sports Med Phys Fitness. 2017;57(10):1261-1266.

5.Roe G, Darrall-Jones J, Till K, et al. Between-day reliability and sensitivity of common fatigue measures in rugby players. Int J Sports Physiol Perform. 2016;11(5):581-586

6.Tavares F, Healey P, Smith TB, Driller M. The effect of training load on neuromuscular performance, muscle soreness and wellness during an in-season non-competitive week in elite rugby athletes. J Sports Med Phys. 2018;58(11):1565-1571

7.Buchheit M, Simpson MB, Al Haddad H, Bourdon PC, Mendez-Villanueva A. Monitoring changes in physical performance with heart rate measures in young soccer players. Eur J App Physiol. 2012;112(2):711-723.

8.Carling C, Lacome M, McCall A, et al. Monitoring of post-match fatigue in professional soccer: welcome to the real world. Sports Med .2018;48(12):2695-2702.

9.Wehbe G, Gabett TJ, Dwyer D, McLellan C, Coad S. Monitoring neuromuscular fatigue in team-sport athletes us-

ing a cycle-ergometer test. Int J Sports Physiol Perform. 2015;10(3):292-297.
10. Wehbe GM, Gabbett TJ, Hartwig TB, McLellan CP . Reliability of a cycle ergometer peak power test in running-based team sport athletes: a technical report. J Strength Cond Res. 2015; 29(7):2050-2055.
11. Ripley NJ, Grainger AT, Comfort P. Comparison between a 6 second and 30 second cycle ergometer test for peak power in elite rugby union players. J Aus Strength Cond. 2018;26(3):27-30.
12. Howard SM, Cumming SP, Atkinson M, Malina RM. Biological maturity-associated variance in peak power output and momentum in academy rugby union players. Eur J Sport Sci 2016;16(8):972-980.
13. Hopkins WG. How to interpret changes in an athletic performance test. Sportscience. 2004;8:1-7.
14. Buchheit M, Lefebvre B, Laursen PB, Ahmaidi S. Reliability, usefulness, and validity of the 30–15 intermittent ice test in young elite ice hockey players. J Strength Cond Res. 2011;25(5):1457-1464.
15. Bell W, Cobner DM. Effect of individual time to peak power output on the expression of peak power output in the 30-s Wingate Anaerobic Test. Int J Sports Med. 2007;28(2):135-139.
16. Hopkins WG. New View of Statistics. Available at: Internet Society for Sport Science: www.sportsci.org. Accessed July 18, 2019.
17. Hopkins WG. Linear models and effect magnitudes for research, clinical and practical applications. Sportscience, 2010;14:49-58.
18. Mathieu B, Peeters A, Piscione J, Lacome M. Usefulness of typical tests of short duration maximal effort used to assess players readiness to perform. Sports Perf Sci Rep. 2017;3(1):1-3.
19. Brehm JW, Self EA. The intensity of motivation. Ann Rev Psychol. 1989;40:109-131.
20. Shearer DA, Kilduff LP, Finn C. Measuring recovery in elite rugby players: The brief assessment of mood, endocrine changes, and power. Res Q Exerc Sport. 2015; 86(4):379-386.
21. Roe G, Darrall-Jones J, Till K. To jump or cycle? Monitoring neuromuscular function in rugby union players. Int J Sports Physiol Perform. 2017;12(5): 690-696.