physiologia

MDPI

Article

Agreement Between Same-Day and Separate-Day Assessment of Critical Power, W′, and Peak Oxygen Uptake in Recreational Cyclists and Triathletes

Caleb S. Clay [1], Noah J. Perez [1], Payton E. Miller [1,2] and John W. Farrell III [1,*]

[1] Sport Training and Athlete Readiness Laboratory, Department of Health and Human Performance, Texas State University, San Marcos, TX 78666, USA; csc118@txstate.edu (C.S.C.)
[2] Neuromuscular and Occupational Performance Laboratory, Department of Kinesiology & Sport Management, Texas Tech University, Lubbock, TX 79409, USA
* Correspondence: jwf77@txstate.edu

Abstract

Objective: The purpose of this exploratory study was to examine mean differences, reliability, and agreement between CP and W′ derived from a 3MAT performed after an ICT and values obtained during a standalone 3MAT performed on a separate day. Methods: Ten recreationally trained cyclists and triathletes completed four laboratory visits, including an ICT followed by a 3MAT after 30 min of recovery (same-day condition) and a standalone 3MAT performed on a separate day. CP and W′ were derived from both conditions and compared using paired *t*-tests and Bland–Altman agreement methods. Results: No systematic differences, but limited individual-level agreement were observed between same-day and separate-day conditions for CP (mean difference: 1.72 ± 36.5 W) or W′ (mean difference: 0.99 ± 5.9 kJ), with trivial to small effect sizes. Peak VO_2 values were also not significantly different across testing conditions ($p = 0.483$). However, Bland–Altman analysis revealed wide limits of agreement for both CP (−69.8 to 73.3 W) and W′ (−10.7 to 12.7 kJ), indicating substantial variability at the individual level. Conclusions: Although same-day testing does not introduce systematic bias in CP or W′ estimation, the wide limits of agreement suggest that these protocols are not interchangeable for individual monitoring. Same-day testing may be appropriate for group-level assessments but should be used with caution when applied to individual athlete monitoring.

Keywords: three-minute all-out test; athlete monitoring; endurance performance; incremental exercise test

Check for updates

Academic Editor: Philip J. Atherton

Received: 22 April 2026
Revised: 3 June 2026
Accepted: 5 June 2026
Published: 16 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Assessments of physiological performance metrics are essential for informing the athlete training process and are commonly conducted using laboratory- or field-based testing [1]. However, comprehensive evaluation of key performance indicators often requires multiple testing sessions, which may disrupt training schedules and reduce ecological validity. As such, there is a growing need for time-efficient testing strategies that provide valid and reliable data while minimizing interference with training.

Critical power (CP) and W′ represent key indicators of endurance performance, reflecting the maximal sustainable metabolic rate and the finite capacity for work performed above CP, respectively [2,3]. Recent reviews have reinforced the applied value of CP and W′ for cycling performance assessment and training prescription, while also emphasizing

that W′ recovery/reconstitution and protocol selection remain important considerations when interpreting power-duration metrics [4,5]. CP and W′ are commonly derived from the three-minute all-out test (3MAT), whereas VO_2max is assessed using an incremental exercise test (IET) [6–8]. Recent methodological reviews further highlight that CP and W′ estimates may vary depending on testing approach, ergometer protocol, and modeling assumptions, underscoring the need to evaluate reliability and agreement when implementing these assessments in applied settings [5,9]. Current recommendations typically require these assessments to be performed on separate days, increasing the logistical burden of testing. Combining these assessments into a single session without compromising measurement validity would reduce time demands and improve the practicality of performance monitoring.

Previous investigations have explored variations in testing protocols used to estimate CP and W′, including modifications to the 3MAT and comparisons between different testing approaches [10–12]. More recent work has continued to emphasize that 3MAT-derived outcomes can demonstrate meaningful subject-level variability, supporting the need to evaluate not only mean differences but also reliability and agreement when comparing protocols [13]. While these studies generally report similar group-level outcomes across protocols, they have not directly examined the comparability of CP and W′ derived from a 3MAT performed immediately following an incremental exercise test (IET) within the same session versus a standalone 3MAT performed on a separate day. Recent literature has highlighted continued interest in CP/W′ assessment, W′ reconstitution, and power-duration modeling, but the specific question of whether same-day IET + 3MAT testing provides acceptable individual-level agreement with separate-day 3MAT assessment remains insufficiently addressed [4,5,13]. As a result, it remains unclear whether same-day testing protocols yield values that are truly interchangeable at the individual level.

Accordingly, it is necessary to determine whether CP and W′ derived from same-day testing are sufficiently consistent with separate-day assessments to support interchangeable use, particularly for individual athlete monitoring. The current study builds on previous work by directly comparing CP, W′, and VO_2max derived from same-day and separate-day testing conditions, with a specific focus on agreement between protocols. Thus, the purpose of this study was to compare CP and W′ values derived from 3MATs performed following an IET under same-day and separate-day conditions. It was hypothesized that no systematic differences would be observed between conditions, but that agreement between protocols may be limited. This distinction between group-level comparability and individual-level agreement is critical for practitioners using these metrics to inform individualized training decisions.

2. Materials and Methods

The current investigation utilized a repeated-measures design to compare physiological outcomes derived from same-day and separate-day testing protocols. Participants completed four laboratory visits. During testing session 1 (TS1), participants performed an incremental cycling test (ICT) to determine maximal oxygen uptake (VO_2max) and power at VO_2max (pVO_2max), followed by a three-minute all-out test (3MAT) after 30 min of recovery (same-day condition). During testing session 2 (TS2), participants performed a standalone 3MAT (separate-day condition). Testing sessions 3 and 4 involved high-intensity intermittent exercise protocols and are reported elsewhere. Critical power (CP) and W′ were derived from both 3MATs and compared between conditions.

2.1. Participants

Recreationally trained cyclists and triathletes were recruited for the current exploratory investigation. Participants were included if they regularly participated in cycling or triathlon training and were familiar with high-intensity cycling exercise. Formal competitive category, weekly training volume, and recent race performance were not collected; therefore, participant training status should be interpreted cautiously. All participants completed a phone screening questionnaire, medical history questionnaire, and Physical Activity Readiness Questionnaire + (PARQ+) to determine that participants were free from cardiovascular, pulmonary, neuromuscular or neurological diseases or conditions [14]. Additionally, participants were not currently taking any cardiovascular or respiratory medications. Prior to exercise testing on TS 1, all participants were confirmed to have normotensive blood pressure. This investigation was approved by the Institutional Review Board of Texas State University (IRB #8340).

2.2. Procedures

Pretesting Guidelines. Participants were instructed to avoid strenuous exercise 48 h before each TS, not to change their diet while enrolled in the study, and to be at least 2 h postprandial prior to visiting the laboratory. All sessions were separated by at least 48 h and no more than 7 days.

Incremental Cycling Test and Verification Protocol. Prior to exercise testing, height and weight were assessed and the participants completed the physical activity rating scale (PA-R). The PA-R asks participants to rate their overall average physical activity levels for the previous 6 months on a scale of 0 (avoid walking or exertion) to 15 (vigorous activity: run 50 miles or more per week or spend 13 or more per week in a comparable physical activity) [15]. The PA-R rating, along with the participant age, sex, and body mass index (BMI) were inserted into the following equations below to first estimate VO_2max then the peak power output (PPO) [16]. Finally, the estimated PPO was then divided by the desired number of stages (10) for the ICT to last to determine both the initial power output and the increments for each stage [16].

$$\text{Estimated } VO_2\text{max} = 56.3 + (1.921 \times \text{PA-R}) - (0.381 \times \text{Age}) - (0.754 \times \text{BMI}) + (10.987 \times \text{Sex})$$

$$(\text{Sex, 1 = Male, 0 = Female})$$

$$\text{PPO} = ((\text{Estimated } VO_2\text{max} - 7) \times (\text{Body Mass})/1.8))/6.12$$

$$\text{Power Output Increments} = \text{Estimated Peak Work Rate}/\text{Desired Number of Stages}$$

$$(\text{Desired Number of Stages} = 10)$$

The cycling ergometer was adjusted to meet each participant's desired specifications, with forward, aft, and height settings for the handlebars and seat recorded for consistency between each TS. The IET began at the predetermined initial power output and increased in power output every minute. Test termination occurred when cadence decreased by >10 RPM from the self-selected cadence for >5 s despite verbal encouragement. Participants were fitted with a chest strap heart rate monitor (Polar, Electro Oy, Kempele, Finland) for continuous monitoring during exercise. Inspired and expired gases were collected (7450 Series Silicone V2 Oro-Nasal Mask, Hans Rudolph, Shawnee, KS, USA) and analyzed via a calibrated open circuit spirometry system. Gas analyzers were calibrated with room air and gases of known concentration before all testing sessions. Upon reaching volitional exhaustion during the IET, the participants immediately completed a 3-min bout of active recovery at 50 W. During this time peak power output (PPO) and pVO_2max were quickly calculated as:

$$\text{PPO} = \text{Wcom} + (t/60 \times \text{POI})$$

where Wcom is the power output, in W, for the final fully completed stage; t is the time in seconds, that the final uncompleted stage was sustained; 60 is the duration, in seconds, for each stage; and POI is the power output increment in W [17]. To account for the lag in the VO_2 kinetics during the ICT, two-thirds the ramp rate was deducted from the PPO to determine pVO_2peak [18–22]:

$$pVO_2max = PPO - (2/3 \times POI)$$

At the end of the 3 min of active recovery, the square-wave VO_2peak verification protocol began. The power output was immediately increased to 105% of the pVO_2max and participants cycled at the same self-selected cadence as during the ICT until reaching volitional exhaustion, >10 RPM decrease from self-selected cadence for >5 s [21,23–25]. Participants then immediately began a 5-min cooldown. Heart Rate Max (HRmax) was defined as the highest 1 s value recorded. VO_2max was determined as the highest 20-s average observed during either the ICT or verification protocol. A supramaximal verification protocol was included because traditional secondary criteria such as heart rate, respiratory exchange ratio, and rating of perceived exertion may not consistently confirm attainment of VO_2max, whereas verification testing provides an additional confirmatory assessment of whether a maximal oxygen uptake has been elicited [8,21,26,27].

Three-Minute All-Out Test. The 3MAT has been shown to be a valid and reliable assessment of both CP and W′ [6,28,29]. After completing 30 min of rest (5-min cool down after verification protocol plus 25 min of passive recovery) following the ICT, participants completed the 3MAT. The fixed resistance for the 3MAT was determined through a series of computations outlined by Dicks et al. 2016 [30]. The gas exchange threshold (GET) was estimated based on a percentage of PPO, determined during the ICT, from normative data, with participants reporting a PA-R of <8 and ≥8 having a GET of 60% and 65% of PPO, respectively [30]. Next, 50% of the difference between GET and PPO was calculated and then divided by 9.8 to determine torque for the 3MAT [30]. Finally, the linear factor was determined by dividing 3MAT torque by participants' body mass, then dividing by 9.8. The 3MAT began with three minutes of unloaded cycling serving as a warmup, with participants instructed to increase their cadence to 110 to 120 revolutions per minute in the final 5 s of the warmup prior to the load setting in [6,31]. Participants were instructed to pedal with an all-out effort and no pacing strategy for the full three minutes. Power data were collected at 4 Hz and exported to Microsoft Excel. CP was determined as the mean power output over the final 30 s of the test, with W′ calculated as the work performed above CP during the first 150 s of the test using the following equation:

$$W' = (150 \times (p150 - CP))/1000)$$

where p150 is the average power output over the first 150 s of the assessment [30,31]. Additional variables calculated included maximum power output (pMax), determined as highest power output achieved during the 3MAT, the average power output over the first (p1Min), second (p2Min), and third (p3Min) minute of the test. Fatigue index (FI) was calculated as the ratio of CP to pMax expressed as a percentage. Also, inspired and expired gases were collected during the 3MAT using the same equipment and procedures as during the ICT, with VO_2peak being defined as the highest 20-s average obtained [6]. The 3MAT was completed on both TS 1 and TS 2 following the same procedures. The only difference between sessions was that TS1 included the ICT and 30 min of recovery before the 3MAT, whereas TS2 included only the standalone 3MAT. Outcomes derived from the 3MAT performed during TS1 were labeled as same-day values, whereas outcomes derived from the standalone 3MAT performed during TS2 were labeled as separate-day values.

Statistical Analysis. All analyses were performed using IBM SPSS Statistics (version 26.0; IBM Corporation, Armonk, NY, USA). Descriptive statistics were used to summarize demographic data. Paired-samples *t*-tests were used to detect differences in CP, W′, pMax, p150, p1min, p2min, p3min, and FI obtained from the two 3MAT conditions (same-day vs. separate-day). A one-way repeated-measures analysis of variance was used to detect differences in maximal/peak VO_2 values recorded during the ICT and the two 3MAT conditions. Agreement between CP and W′ measured during the two 3MAT conditions was assessed using mean difference, 95% confidence intervals, and Bland–Altman limits of agreement. Reliability and measurement error were further assessed using intraclass correlation coefficients (ICC), typical error, standard error of measurement (SEM), and coefficient of variation (CV). ICCs were calculated using an absolute-agreement model to evaluate the consistency of same-day and separate-day values. Typical error was calculated as the standard deviation of the differences divided by $\sqrt{2}$, and SEM was used to quantify absolute measurement error. CV was calculated as typical error divided by the mean of the paired measurements and expressed as a percentage. Pearson correlation coefficients were calculated as secondary descriptive indicators of association between CP and W′ values assessed during the two 3MAT conditions and were not interpreted as evidence of agreement. Partial eta squared (η^2) was used to interpret effect sizes for the ANOVA, with values of 0.02, 0.13, and 0.26 indicative of small, medium, and large effects, respectively [32]. Effect sizes for pairwise comparisons were interpreted using Cohen's d, with values of <0.2, 0.2 to 0.6, >0.6 to 1.2, >1.2 to 2.0, and >2.0 to 4.0 classified as trivial, small, moderate, large, and very large, respectively [33]. Statistical significance was set at $p < 0.05$, with data reported as mean ± standard deviation. Given the exploratory nature of the study and fixed sample size, a sensitivity analysis was conducted to estimate the smallest detectable paired effect size at $\alpha = 0.05$ and 80% power. With $n = 10$, the study was powered to detect large paired effects of approximately dz = 1.00, corresponding to approximately 36.4 W for CP and 6.0 kJ for W′. Therefore, the study was not powered to detect small-to-moderate differences between protocols.

3. Results

Nine recreationally trained cyclists and one recreationally trained triathlete ($n = 10$; 8 males, mean ± SD: age 26.5 ± 5.6 yr, height 171.2 ± 5.82 cm, body mass 69.8 ± 11.4 kg) were included in the current analysis. During the ICT, participants achieved a PPO of 339.1 ± 55.8 W and a VO_2max of 55.9 ± 6.8 mL·kg^{-1}·min^{-1}. Paired samples *t*-tests revealed no significant differences between same-day and separate-day 3MAT conditions for any performance variables, with trivial to small effect sizes observed (Table 1).

Table 1. Performance metrics from the two 3MAT conditions ($n = 10$).

	Same-Day 3MAT		Separate-Day 3MAT			
	Mean ± SD	**Range**	**Mean ± SD**	**Range**	***p***	***d***
Variable						
VO_2peak (mL·kg^{-1}·min^{-1})	53.6 ± 5.9	43.17–61.1	52.6 ± 6.4	48.0–59.7		
pMax (W)	708.1 ± 166.2	500.0–1077.0	775.0 ± 182.0	516.0–1042.0	0.08	0.38
p150 (W)	350.2 ± 56.9	230.0–461.2	355.8 ± 58.6	238.7–466.8	0.27	0.09
CP (W)	267.3 ± 34.7	185.7–299.7	269.0 ± 48.2	173.2–339.7	0.88	0.04
W′ (kJ)	12.4 ± 5.8	6.7–27.1	13.4 ± 5.4	8.4–26.2	0.61	0.13
p1min (W)	449.0 ± 75.0	297.0–582.7	459.8 ± 79.9	312.7–603.2	0.20	0.04
p2min (W)	288.7 ± 56.1	186.5–401.7	298.2 ± 50.6	193.5–387.4	0.10	0.06

Table 1. *Cont.*

	Same-Day 3MAT		Separate-Day 3MAT			
	Mean ± SD	**Range**	**Mean ± SD**	**Range**	***p***	***d***
Variable						
p3min (W)	271.4 ± 36.4	184.5–308.7	272.5 ± 47.7	175.1–345.9	0.89	0.02
FI (%)	61.2 ± 7.8	51.1–73.9	63.5 ± 11.0	43.9–77.81	0.40	0.24

Abbreviations: $3MAT_{same\text{-}day}$ = three-minute all-out test performed on same day as maximal oxygen consumption test; $3MAT_{separate\text{-}day}$ = three-minute all-out test performed on separate day as maximal oxygen consumption test; VO_2peak = peak oxygen consumption; pMax = maximal power output; p150 = average power output over first 150 s; CP = critical power; W′ = W prime; p1min = average power output during the first minute; p2min = average power output during the second minute; p3min = average power output during the third minute; FI = fatigue index; *d* = Cohen's d effect sizes.

Similarly, no significant differences were observed among peak VO_2 values obtained during the ICT/verification and the two 3MAT conditions using a one-way repeated-measures ANOVA (F = 0.749, p = 0.483, η^2 = 0.05). Only nine participants were included in this analysis because gas collection was discontinued for one participant due to discomfort with the mask. Bland–Altman analyses for CP and W′ are presented in Figures 1 and 2. The mean difference for CP was 1.72 ± 36.5 W (95% CI: −24.4 to 27.8 W), with limits of agreement ranging from −69.8 to 73.3 W. Additionally, the ICC for absolute agreement was 0.646, typical error was 25.8 W, SEM was 25.0 W, and CV was 9.6% for CP. For W′, the mean difference was 0.99 ± 5.9 kJ (95% CI: −3.3 to 5.2 kJ), with limits of agreement ranging from −10.7 to 12.7 kJ. Also for W′, the ICC for absolute agreement was 0.445, typical error was 4.2 kJ, SEM was 4.2 kJ, and CV was 32.8%. Pearson correlation analysis revealed a moderate association between CP measured under same-day and separate-day conditions (r = 0.656, p = 0.039), but no significant association was observed for W′ (r = 0.428, p = 0.217).

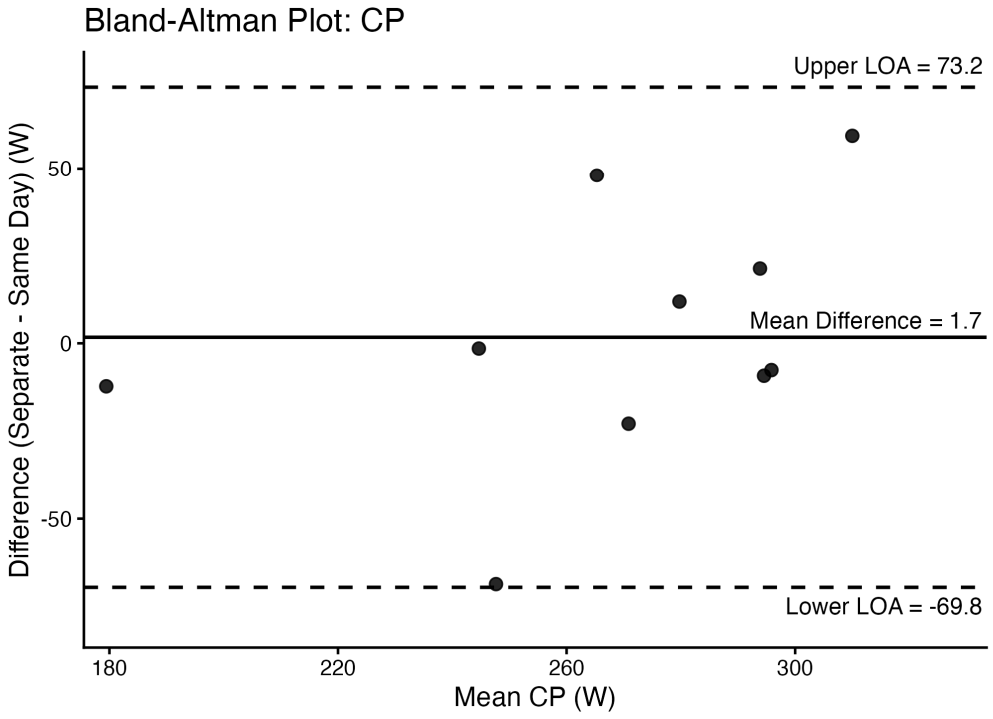


Figure 1. Bland–Altman plot illustrating agreement between same-day and separate-day critical power (CP) measurements. The solid line represents the mean difference (1.7 W), and dashed lines represent the upper (73.3 W) and lower (−69.8 W) limits of agreement.

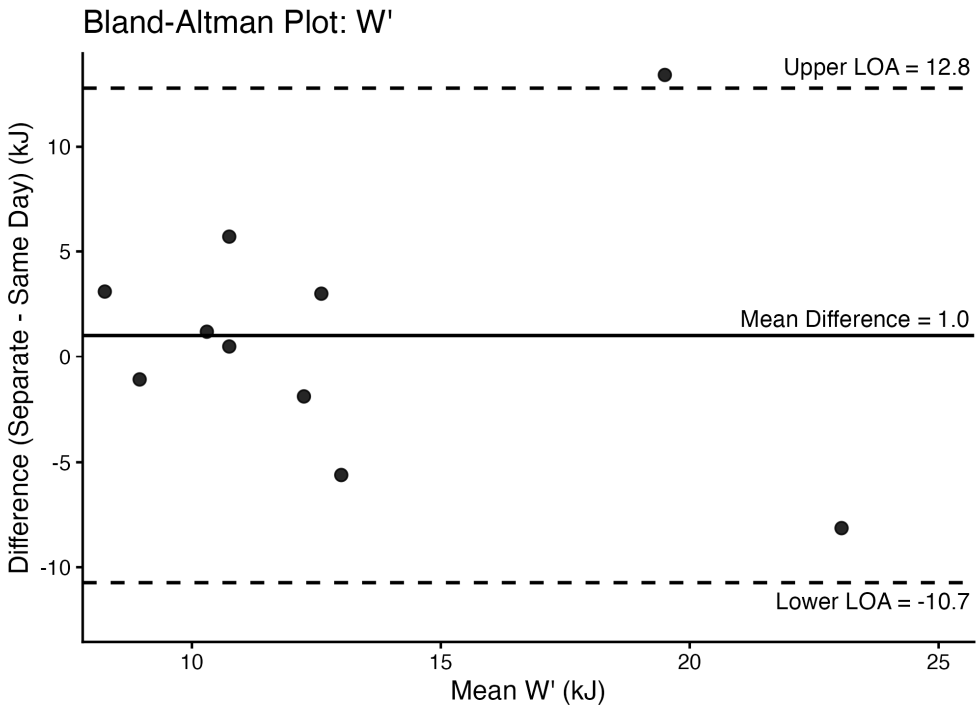


Figure 2. Bland–Altman plot illustrating agreement between same-day and separate-day W′ measurements. The solid line represents the mean difference (1.0 kJ), and dashed lines represent the upper (12.7 kJ) and lower (−10.7 kJ) limits of agreement.

4. Discussion

The purpose of this study was to determine whether critical power (CP) and W′ derived from a 3MAT performed following an ICT are comparable to values obtained from a standalone 3MAT performed on a separate day. No systematic differences were observed between same-day and separate-day conditions for CP (mean difference: 1.72 ± 36.5 W) or W′ (mean difference: 0.99 ± 5.9 kJ), with trivial to small effect sizes. However, Bland–Altman analysis revealed wide limits of agreement for both CP (−69.8 to 73.3 W) and W′ (−10.7 to 12.7 kJ), indicating substantial variability at the individual level. Notably, the limits of agreement for CP represented a large range of potential error relative to the mean CP value, while the limits of agreement for W′ were large relative to the group mean, further highlighting the lack of interchangeability between protocols at the individual level. A moderate correlation was observed for CP between conditions (r = 0.656), whereas W′ demonstrated weaker, non-significant associations. These findings extend previous investigations that reported similar mean CP and W′ values between combined and isolated testing protocols by demonstrating that, despite comparable group-level outcomes, substantial variability exists at the individual level when agreement is assessed [11]. This interpretation is also consistent with recent work emphasizing that 3MAT-derived outcomes may demonstrate meaningful subject-level variability, reinforcing the need to evaluate reliability and agreement rather than relying exclusively on group mean comparisons [13]. Collectively, these findings indicate that while same-day testing does not introduce systematic bias, it lacks sufficient agreement for interchangeable use in individual athlete monitoring.

Critical power (CP) represents a key metabolic threshold with strong implications for endurance performance, while W′ reflects the finite capacity for work performed above CP [2]. Recent reviews continue to support the applied value of CP and W′ for endurance performance assessment, cycling power profiling, and training prescription, while emphasizing that protocol selection and W′ reconstitution can influence interpretation

of these metrics [4,5]. Together with maximal oxygen uptake (VO_2max), these metrics provide a comprehensive profile of an individual's physiological capacity during endurance exercise [2]. However, current testing recommendations require CP/W′ and VO_2max to be assessed on separate days, increasing logistical burden and disrupting training schedules [6,31]. As a result, there is growing interest in developing time-efficient testing strategies that minimize disruption while maintaining measurement validity. Previous investigations have explored single-session approaches to estimate CP using modified 3MAT protocols, demonstrating the potential feasibility of one-day testing [11,12]. The present findings build on this work by demonstrating that, although same-day testing may provide comparable group-level estimates, the substantial variability observed at the individual level limits its practical utility for athlete monitoring. As such, while a single-session approach may reduce testing burden, its implementation should be approached with caution when precise individual assessment is required. These findings highlight an important distinction between group-level validity and individual-level agreement, which is critical when interpreting performance metrics in applied settings.

The absence of systematic differences in CP between conditions is consistent with previous work that also observed similar CP values between combined and isolated testing protocols [11]. However, in contrast to the narrower limits of agreement reported in that study, the present investigation demonstrated substantially wider limits of agreement, indicating greater variability at the individual level. Several methodological differences may explain this discrepancy. Constantini et al. implemented a shorter recovery period of 20 min, whereas the present study used a 30-min recovery period [11]. Although the present study did not directly measure recovery kinetics, lactate, phosphocreatine resynthesis, or muscle metabolic responses, differences in recovery duration may have contributed to individual variability prior to the 3MAT. A longer recovery period could permit substantial but individually variable recovery following maximal exercise, which may influence subsequent high-intensity performance and contribute to wider limits of agreement [34]. This interpretation remains speculative because the present study did not include direct measures of fatigue, lactate, phosphocreatine recovery, metabolite accumulation, muscle metabolic responses, or training load between sessions. Therefore, the mechanisms underlying the wider limits of agreement cannot be determined from the present data and should be tested directly in future work using metabolic, perceptual, and workload-recovery markers.

W′ represents the finite capacity to perform work above CP and is thought to reflect the integrated contribution of anaerobic energy provision, intramuscular phosphocreatine availability, metabolite accumulation, and fatigue-related limitations during severe-intensity exercise [1,2]. Consistent with previous investigations, no systematic differences in W′ were observed between combined and isolated testing conditions, suggesting that same-day testing does not introduce a consistent directional effect at the group level [11]. However, similar to CP, wide limits of agreement were observed, and in contrast to CP, no significant correlation was detected between conditions. The greater variability observed for W′ is physiologically plausible given that W′ appears more sensitive than CP to prior exercise and recovery state. Prior exhaustive exercise can reduce W′ while leaving CP relatively unchanged, suggesting that W′ may be more affected by residual fatigue or incomplete recovery [35,36]. Mechanistically, prior severe-intensity exercise may reduce the capacity to perform subsequent work above CP through incomplete phosphocreatine restoration, accumulation or incomplete clearance of fatigue-related metabolites, altered muscle contractile function, and individual differences in W′ reconstitution kinetics. Recent work on W′ recovery and reconstitution further supports the concept that W′ recovery is dynamic, intensity-dependent, and individually variable [4,36,37]. Although the present study did

not directly assess phosphocreatine recovery, metabolite accumulation, lactate, or perceived exertion, these mechanisms provide a reasonable physiological context for interpreting the poor agreement observed for W′. Future studies should include direct metabolic and perceptual markers to determine whether these factors explain individual variability in same-day testing responses.

No systematic differences were observed among peak VO_2 values obtained during the ICT/verification protocol and the two 3MAT conditions, suggesting that, at the group level, peak oxygen uptake responses were similar across the ICT and 3MAT conditions in the present sample. These findings are consistent with previous work demonstrating stable VO_2peak values across varying recovery durations and protocols incorporating multiple high-intensity efforts within a single testing session [10]. However, VO_2peak obtained during a 3MAT should be interpreted cautiously because the structure, pacing characteristics, and physiological demands of a 3MAT differ from those of a traditional incremental or verification protocol. Therefore, the present findings should not be interpreted as evidence that the ICT and 3MAT are interchangeable methods for VO_2peak assessment, but rather that prior ICT exposure did not appear to substantially alter group-level VO_2peak responses during the subsequent 3MAT. Although prior literature suggests that VO_2peak may be relatively stable across repeated maximal efforts, the present study did not include direct measures of fatigue, lactate, substrate utilization, or recovery status [8,26]. Future studies should include these markers to better determine whether accumulated fatigue or recovery state influences VO_2 responses during same-day testing.

The ability to assess CP and W′ within a single testing session would provide a more practical approach for coaches, researchers, and sport scientists by reducing logistical burden and minimizing disruptions to training. Previous investigations have demonstrated that single-day testing can produce comparable group-level outcomes, and the present findings support this observation [11,12]. However, the wide limits of agreement observed in the present study indicate that same-day and separate-day protocols should not be considered interchangeable for individual monitoring. This distinction is important because practitioners often use CP and W′ to guide training prescription, evaluate athlete readiness, and monitor longitudinal changes in performance capacity. In the present study, the limits of agreement for CP were approximately ±70 W, which represents a large range of potential error relative to the mean CP value and could meaningfully alter training-zone classification. Similarly, the limits of agreement for W′ were approximately ±11–13 kJ, which is large relative to the group mean and could obscure meaningful changes in high-intensity work capacity. Therefore, although same-day testing may be acceptable for evaluating group-level trends when logistical constraints prevent multiple testing sessions, separate-day testing remains preferable when CP and W′ are used to make individual training or monitoring decisions.

The present study is not without limitations. The small sample size and predominance of male participants limit the generalizability of these findings. Additionally, only a single recovery duration (30 min) was examined, and different recovery intervals may influence the level of agreement between testing conditions. While no systematic differences were observed between same-day and separate-day protocols, the wide limits of agreement substantially reduce the practical applicability of same-day testing for individual performance monitoring. The small sample size limits the precision of the Bland–Altman limits of agreement and reduces confidence in the stability of reliability estimates. Therefore, the present findings should be interpreted as preliminary and hypothesis-generating rather than definitive evidence regarding protocol interchangeability. An additional limitation is that the present study did not include direct markers of fatigue, recovery, or energetic status, such as blood lactate, phosphocreatine recovery, metabolite accumulation, sub-

strate utilization, RPE, hydration, sleep, or training load between sessions. As a result, the mechanisms underlying the observed variability in CP and W′ agreement cannot be determined directly from these data. The physiological explanations should therefore be interpreted as plausible mechanisms based on prior literature rather than mechanisms confirmed by the present study. Although participants were instructed to avoid strenuous exercise for 48 h, maintain their normal diet, and arrive at least 2 h postprandial, caffeine intake, hydration status, sleep duration/quality, and training load between sessions were not objectively monitored. These factors may have influenced performance outcomes, particularly given the small sample size. Future research should expand the participant pool to include a broader range of training statuses, ages, and female athletes, while also examining the influence of varying recovery durations on the reliability of CP and W′ assessments within a single session. Collectively, these findings reinforce the need for caution when implementing same-day testing protocols in applied settings.

5. Conclusions

This exploratory study found no systematic mean differences in CP or W′ between same-day and separate-day 3MAT protocols; however, reliability indices and wide limits of agreement indicate limited individual-level interchangeability, particularly for W′. Peak VO_2 responses did not differ significantly across the ICT/verification and 3MAT conditions, although VO_2peak obtained during a 3MAT should not be interpreted as methodologically equivalent to VO_2max obtained during a traditional incremental test with verification. In line with recent work emphasizing protocol-dependent variability in CP and W′ estimation, subject-level variability in 3MAT outcomes, and individual differences in W′ reconstitution, the present findings suggest that same-day testing may be useful when logistical constraints require group-level assessment. However, separate-day testing remains preferable when CP and W′ are used for individual athlete monitoring, training prescription, or longitudinal performance evaluation.

Author Contributions: Conceptualization, N.J.P., P.E.M. and J.W.F.III; methodology, J.W.F.III; formal analysis, N.J.P., P.E.M. and J.W.F.III; investigation, N.J.P., P.E.M. and J.W.F.III; writing—original draft preparation, C.S.C. and J.W.F.III; writing—review and editing, N.J.P., P.E.M. and J.W.F.III; supervision, J.W.F.III. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Texas State University (IRB #8340 approve date: 30 June 2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available upon reasonable request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jones, A.M.; Vanhatalo, A. The 'Critical Power' Concept: Applications to Sports Performance with a Focus on Intermittent High-Intensity Exercise. *Sports Med.* **2017**, *47*, 65–78. [CrossRef] [PubMed]
2. Poole, D.C.; Burnley, M.; Vanhatalo, A.; Rossiter, H.B.; Jones, A.M. Critical Power: An Important Fatigue Threshold in Exercise Physiology. *Med. Sci. Sports Exerc.* **2016**, *48*, 2320–2334. [CrossRef] [PubMed]
3. Jones, A.M.; Vanhatalo, A.; Burnley, M.; Morton, R.H.; Poole, D.C. Critical Power: Implications for Determination of $\dot{V}O_{2max}$ and Exercise Tolerance. *Med. Sci. Sports Exerc.* **2010**, *42*, 1876–1890. [CrossRef] [PubMed]
4. Chorley, A.; Lamb, K.L. The Application of Critical Power, the Work Capacity above Critical Power (W′), and Its Reconstitution: A Narrative Review of Current Evidence and Implications for Cycling Training Prescription. *Sports* **2020**, *8*, 123. [CrossRef]

5. Leo, P.; Spragg, J.; Podlogar, T.; Lawley, J.S.; Mujika, I. Power Profiling and the Power-Duration Relationship in Cycling: A Narrative Review. *Eur. J. Appl. Physiol.* **2022**, *122*, 301–316. [CrossRef]
6. Burnley, M.; Doust, J.H.; Vanhatalo, A. A 3-Min All-Out Test to Determine Peak Oxygen Uptake and the Maximal Steady State. *Med. Sci. Sports Exerc.* **2006**, *38*, 1995–2003. [CrossRef]
7. Vanhatalo, A.; Jones, A.M.; Burnley, M. Application of Critical Power in Sport. *Int. J. Sports Physiol. Perform.* **2011**, *6*, 128–136. [CrossRef]
8. Poole, D.C.; Jones, A.M. Measurement of the Maximum Oxygen Uptake VO_{2max}: VO_{2peak} Is No Longer Acceptable. *J. Appl. Physiol.* **2017**, *122*, 997–1002. [CrossRef]
9. Lipková, L.; Kumstát, M.; Struhár, I. Determination of Critical Power Using Different Possible Approaches among Endurance Athletes: A Review. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7589. [CrossRef]
10. Karsten, B.; Hopker, J.; Jobson, S.A.; Baker, J.; Petrigna, L.; Klose, A.; Beedie, C. Comparison of Inter-Trial Recovery Times for the Determination of Critical Power and W' in Cycling. *J. Sports Sci.* **2017**, *35*, 1420–1425. [CrossRef]
11. Constantini, K.; Sabapathy, S.; Cross, T.J. A Single-Session Testing Protocol to Determine Critical Power and W′. *Eur. J. Appl. Physiol.* **2014**, *114*, 1153–1161. [CrossRef] [PubMed]
12. Bergstrom, H.C.; Housh, T.J.; Zuniga, J.M.; Camic, C.L.; Traylor, D.A.; Schmidt, R.J.; Johnson, G.O. A New Single Work Bout Test to Estimate Critical Power and Anaerobic Work Capacity. *J. Strength Cond. Res.* **2012**, *26*, 656–663. [CrossRef]
13. Sreedhara, V.S.M.; Mocko, G.M.; Hutchison, R.E. Repeatability and Variability of the 3-Min All-Out Test at the Subject Level. *J. Sci. Sport Exerc.* **2023**, *5*, 77–86. [CrossRef]
14. Warburton, D.E.R.; Bredin, S.S.D.; Jamnik, V.K.; Gledhill, N. International Launch Of The Par Q+ And ePARmed X+ Validation of the PAR Q+ and ePARmed X+. *Health Fit. J. Can.* **2011**, *4*, 38–46.
15. George, J.D.; Stone, W.J.; Burkett, L.N. Non-Exercise $\dot{V}O_{2max}$ Estimation for Physically Active College Students. *Med. Sci. Sports Exerc.* **1997**, *29*, 415–423. [CrossRef]
16. Jamnick, N.A.; By, S.; Pettitt, C.D.; Pettitt, R.W. Comparison of the YMCA and a Custom Submaximal Exercise Test for Determining $V\dot{O}_{2max}$. *Med. Sci. Sports Exerc.* **2016**, *48*, 254–259. [CrossRef] [PubMed]
17. Luttikholt, H.; McNaughton, L.R.; Midgley, A.W.; Bentley, D.J. A Prediction Model for Peak Power Output From Different Incremental Exercise Tests. *Int. J. Sports Physiol. Perform.* **2006**, *1*, 122–136. [CrossRef]
18. Whipp, B.J.; Davis, J.A.; Torres, F.; Wasserman, K. A Test to Determine Parameters of Aerobic Function during Exercise. *J. Appl. Physiol.* **1981**, *50*, 217–221. [CrossRef]
19. Beaver, W.L.; Wasserman, K.; Whipp, B.J. A New Method for Detecting Anaerobic Threshold by Gas Exchange. *J. Appl. Physiol.* **1986**, *60*, 2020–2027. [CrossRef]
20. Jones, A.M.; Berger, N.J.A.; Wilkerson, D.P.; Roberts, C.L. Effects of "Priming" Exercise on Pulmonary O_2 Uptake and Muscle Deoxygenation Kinetics during Heavy-Intensity Cycle Exercise in the Supine and Upright Positions. *J. Appl. Physiol.* **2006**, *101*, 1432–1441. [CrossRef]
21. Rossiter, H.B.; Kowalchuk, J.M.; Whipp, B.J. A Test to Establish Maximum O_2 Uptake despite No Plateau in the O_2 Uptake Response to Ramp Incremental Exercise. *J. Appl. Physiol.* **2006**, *100*, 764–770. [CrossRef]
22. Hill, D.W.; Poole, D.C.; Smith, J.C. The Relationship between Power and the Time to Achieve VO_{2max}. *Med. Sci. Sports Exerc.* **2002**, *34*, 709–714.
23. Midgley, A.W.; McNaughton, L.R.; Carroll, S. Verification Phase as a Useful Tool in the Determination of the Maximal Oxygen Uptake of Distance Runners. *Appl. Physiol. Nutr. Metab.* **2006**, *31*, 541–548. [CrossRef]
24. Beltz, N.M.; Gibson, A.L.; Janot, J.M.; Kravitz, L.; Mermier, C.M.; Dalleck, L.C. Graded Exercise Testing Protocols for the Determination of VO_2 Max: Historical Perspectives, Progress, and Future Considerations. *J. Sports Med.* **2016**, *2016*, 3968393. [CrossRef]
25. Foster, C.; Kuffel, E.; Bradley, N.; Battista, R.A.; Wright, G.; Porcari, J.P.; Lucia, A.; de Koning, J.J. VO_{2max} during Successive Maximal Efforts. *Eur. J. Appl. Physiol.* **2007**, *102*, 67–72. [CrossRef] [PubMed]
26. Midgley, A.W.; McNaughton, L.R.; Polman, R.; Marchant, D. Criteria for Determination of Maximal Oxygen Uptake: A Brief Critique and Recommendations for Future Research. *Sports Med.* **2007**, *37*, 1019–1028. [CrossRef] [PubMed]
27. Costa, V.A.B.; Midgley, A.W.; Carroll, S.; Astorino, T.A.; De Paula, T.; Farinatti, P.; Cunha, F.A. Is a Verification Phase Useful for Confirming Maximal Oxygen Uptake in Apparently Healthy Adults? A Systematic Review and Meta-Analysis. *PLoS ONE* **2021**, *16*, e0247057. [CrossRef] [PubMed]
28. Black, M.I.; Durant, J.; Jones, A.M.; Vanhatalo, A. Critical Power Derived from a 3-min All-out Test Predicts 16.1-km Road Time-trial Performance. *Eur. J. Sport Sci.* **2014**, *14*, 217–223. [CrossRef]
29. Clark, I.E.; West, B.M.; Reynolds, S.K.; Murray, S.R.; Pettitt, R.W. Applying the Critical Velocity Model for an Off-Season Interval Training Program. *J. Strength Cond. Res.* **2013**, *27*, 3335–3341. [CrossRef]
30. Dicks, N.D.; Jamnick, N.A.; Murray, S.R.; Pettitt, R.W. Load Determination for the 3-Minute All-Out Exercise Test for Cycle Ergometry. *Int. J. Sports Physiol. Perform.* **2016**, *11*, 197–203. [CrossRef]

31. Vanhatalo, A.; Doust, J.H.; Burnley, M. Determination of Critical Power Using a 3-Min All-out Cycling Test. *Med. Sci. Sports Exerc.* **2007**, *39*, 548–555. [CrossRef]
32. Fatma Hilal, Y.; Abdulvahap, P.; Georgian, B. Statistical Effect Sizes in Sports Science. *J. Exerc. Sci. Phys. Act. Rev.* **2024**, *2*, 164–1710. [CrossRef]
33. Hopkins, W.G.; Marshall, S.W.; Batterham, A.M.; Hanin, J. Progressive Statistics for Studies in Sports Medicine and Exercise Science. *Med. Sci. Sports Exerc.* **2009**, *41*, 3–12. [CrossRef] [PubMed]
34. Forbes, S.C.; Raymer, G.H.; Kowalchuk, J.M.; Thompson, R.T.; Marsh, G.D. Effects of Recovery Time on Phosphocreatine Kinetics during Repeated Bouts of Heavy-Intensity Exercise. *Eur. J. Appl. Physiol.* **2008**, *103*, 665–675. [CrossRef]
35. Ferguson, C.; Rossiter, H.B.; Whipp, B.J.; Cathcart, A.J.; Murgatroyd, S.R.; Ward, S.A. Effect of Recovery Duration from Prior Exhaustive Exercise on the Parameters of the Power-Duration Relationship. *J. Appl. Physiol.* **2010**, *108*, 866–874. [CrossRef] [PubMed]
36. Welburn, A.J.; Pugh, C.F.; Bailey, S.J.; Ferguson, R.A. W′ Reconstitution Modelling during Intermittent Exercise Performed to Task Failure. *Eur. J. Appl. Physiol.* **2026**, *126*, 765–778. [CrossRef]
37. Chorley, A.; Bott, R.P.; Marwood, S.; Lamb, K.L. Bi-Exponential Modelling of W′ Reconstitution Kinetics in Trained Cyclists. *Eur. J. Appl. Physiol.* **2022**, *122*, 677–689. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.