AN ABSTRACT OF THE THESIS OF

Lee Arakawa for the degree of Master of Science in Kinesiology presented on June 22, 2018.

Title: Assessing Between-day and Inter-rater Reliability of a 3-minute Running Test for Determining Critical Speed

Abstract approved: ____

Marc F. Norcross

For athletes and training populations, the ability to track progress and determine appropriate workloads to stimulate adaptation is vital. Measures that represent a person's aerobic fitness such as VO_2 max and blood lactate concentration are often used for these purposes. While these measures can reveal a lot about a person's aerobic fitness, they also require invasive methods and/or testing procedures that require a lot of time and money. Critical speed (CS) is an indicator of aerobic fitness that represents the greatest metabolic rate that results in the rate of lactate production being matched by the rate of clearance. It also precisely predicts aerobic performance and tolerance. Testing for critical speed traditionally required multiple bouts to exhaustion, until the 3-minute testing model was developed. Though the validity of a 3-minute running test using minimal equipment has been shown, the purpose of this study was to assess the between-day and inter-rater reliability of this test. Sixteen healthy participants completed an all-out 3-minute running test and served as a rater for one other participant. Eleven participants also completed a second all-out running test 3-14 days after the initial session. From time splits recorded by the participants (Session 1) and an expert rater (Sessions 1 and 2), CS was calculated and

inter-rater and between-day reliability assessed. The primary findings of this study are that the use of the 3-minute running test as applied in this study requires no specialized training as evidenced by the fact that CS values obtained by individuals with minimal training were consistent to those obtained by an expert rater. However, while the test-retest reliability of the 3-minute test was strong-excellent (ICC = 0.82), there was a systematic bias detected in which participants exhibited significantly greater CS values during Session 2 compared to Session 1, and the magnitude of this difference (0.41 m/s) was determined to be clinically meaningful. Therefore, it was concluded that though the test could be conducted consistently by coaches, physical education teachers, or personal trainers with minimal familiarization and equipment, the use of this test for determining critical speed without the application of greater testing controls is not supported due to the fact that the magnitude of the difference in CS between testing sessions was large enough that it would lead to different interpretations of an individual's cardiovascular fitness.

Assessing Between-day and Inter-rater Reliability of a 3-minute Running Test for Determining Critical Speed

by
Lee Arakawa

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented June 22, 2018
Commencement June 2019

Master of Science thesis of Lee Arakawa presented on June 22, 2018

APPROVED:

Major Professor, representing Kinesiology

Head of the School of Biological and Population Health Sciences

---

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Lee Arakawa, Author

ACKNOWLEDGEMENTS

To Dr. Norcross, thank you for your selfless act of kindness that has completely changed the course of my life. You took me on as a student when you didn’t have to and even though I entered through an untraditional route you always made me feel that I belonged. You took a chance on me and my life became so much better for it. Being able to learn from you and work with you gave me the confidence and belief in myself to figure out what I wanted in a career and to pursue it. I hope you know how grateful I truly am for all that you have done for me.

To the members of my committee, Dr. Sam Johnson, Dr. John Schuna Jr., Dr. Karen Elliott, and Dr. Jay Penry, thank you firstly for being a part of my committee. Also, thank you for all your feedback and support in the development of this thesis project.

To my family, even though I am thousands of miles from home I never feel that I am without the love and support that you give me. You all empower me to follow my dreams and be the best that I can be. And a special thank you to Lyndsey and Lilo, you two are my everything. Thank you for your patience, understanding, and support for me during this process and for all the happiness you bring into my life. I love you all.

CONTRIBUTION OF AUTHORS

Dr. Norcross assisted with the conception and design of the study, data analysis and interpretation, and critical revision for important intellectual content. Dr. Penry also assisted with study design.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 – INTRODUCTION

For athletes and populations interested in training and performance, it is valuable to have ways of quantifying their athletic abilities. Without any way to quantify their speed, power, or strength, it would be very difficult to determine whether or not they are ever improving. Such measures allow athletes to track their progress over time and/or determine appropriate workloads – or training stimuli – to be used when creating training programs. The identification of the appropriate training stimulus is important because an insufficient training stimulus can lead to a lack of progress or a detraining effect, while too great of an stimulus can result in an overtraining effect (Kenttä & Hassmén, 1998).

There are specific measures of fitness associated with each of the fitness categories of muscular strength, flexibility, muscular endurance, aerobic fitness, and body composition (Caspersen & Christenson, 1985). Aerobic fitness, or cardiovascular endurance, can be defined as the capacity to sustain a given velocity for the longest possible time (Jones & Carter, 2000). Representing cardiovascular endurance can be done by a velocity-time curve, also represented as a power-duration relationship in other fitness domains (Jones & Carter, 2000). The power-duration relationship describes exercise tolerance within the severe- and extreme-intensity domains and is comprised of the parameters critical power and a work constant (W’) (Hill, 1993).

When considering cardiovascular endurance during running, the same parameters are referred to as critical speed and a distance constant (D’) which is determined by the boundaries of the severe-intensity domain (Solomonson et al., 2016). Critical speed is determined by the highest rate of oxygen utilization matched by oxygen delivery (Skiba et al., 2012), and distinguishes the threshold above which $\dot{V}O_2$, blood flow and intramuscular concentrations of phosphocreatine, inorganic phosphate and hydrogen ions are no longer able to achieve steady-state values and exercise tolerance is predictably limited (Vanhatalo et al., 2010). Therefore, critical speed marks the boundary between the heavy- and severe-intensity domains and has important implications for exercise testing and prescription by helping identify the workload and training stimulus necessary for an optimal training effect (Broxterman et al., 2014).

Traditionally, the testing to determine a person's critical speed required multiple separate bouts to exhaustion conducted across the span of multiple days. To determine a person's critical speed in running, participants would first complete a laboratory-based test for determining their VO_2 max, then they would complete runs to exhaustion at speeds of 90%, 100%, 120%, and 140% of the speed that elicited their initially measured VO_2 max (Di Prampero, 1999). These types of test designs may elicit valid results, but they require a lot of time, and a lot of energy expenditure from the participants.

To address the issue of the test structure taking up too much time, it was determined that an all-out 3-minute running test conducted during a single session could accurately determine a person's critical speed. It was proposed that the 3-minute length is just long enough to yield a stable power output, but short enough such that the subjects could complete the test (Burnley, 2009). Several studies have conducted the 3-minute test on running models and have shown that it is a valid testing method (Broxterman et al., 2013);(Pettitt et al., 2012);(Clark et al., 2013). However, these studies required the use of expensive equipment such as GPS (Pettitt et al., 2012), accelerometers (Broxterman et al., 2013), and camera systems (Clark et al., 2013) to record their measurements.

Recently, Maryn et al. (n.d.) demonstrated the validity of an all-out 3-minute field test for determining critical speed with the use of cones and a stopwatch instead of high-cost testing equipment. However, while these results provide initial evidence in support of a cost and time efficient method to determine critical speed, the reliability of this testing protocol is unknown. If reliable, the Maryn 3-minute test would provide an accurate and consistent test that could significantly alter the way in which exercise intensities are determined and prescribed leading to better training effects and results. It would also add to the practicality of the test if the protocols of the test were shown to be replicable by any practitioner, coach, staff member, or even a fellow athlete. If the test can be conducted by anyone with minimal training, it could potentially lead to more widespread use of the test.

Problem Statement

The *objective* of this study was to examine the reliability of the Maryn 3-minute test (Maryn et al., 2017). The *central hypothesis* was that critical speed as determined using the Maryn 3-minute test would demonstrate adequate reliability for clinical use (intraclass correlation coefficient (ICC>0.7)) in aerobically trained, college-aged men and women.

Specific Aims and Hypotheses

Aim #1. To assess the between-day test-retest reliability of the Maryn 3-minute test.

We hypothesized that critical speed as determined using the Maryn 3-minute test would be consistent (ICC > 0.7) when performed on two different occasions separated by 3 - 14 days.

Aim #2. To assess the inter-rater reliability of the Maryn 3-minute test.

We hypothesized that there would be agreement (ICC > 0.7) in the critical speed values obtained by different raters assessing the same trial of the Maryn 3-minute test.

Assumptions and Limitations

For the Maryn 3-minute test, participants needed to run with maximal effort from start to finish. It would present a limitation within our test if participants paced, or did not run at full speed and maximal effort at any point throughout the 3-minute duration of the test. It was assumed that participants did not participate in strenuous physical activity 24 hours prior to testing, and that there was no effect from diet or training during the duration of the study. We also assumed that there would be no difference in responses to the verbal encouragement between sessions.

Significance

Critical speed precisely predicts exercise performance, determines the feasibility of an athletic feat, and can model optimal performance tactics for a team, or group of athletes, that has different critical speed or D' values (Poole et al., 2016). Testing for critical speed used to be inconvenient in that it required multiple testing sessions (Di Prampero, 1999). This challenge was addressed with the development of the 3-minute test. However, even though the 3-minute test was shown to be a valid way of determining a person's critical speed, the feasibility of using this test was limited by the need for expensive equipment. This limitation was addressed when Maryn modified the 3-minute test so that it could be conducted with the use of low-cost cones and a stopwatch (Maryn et al., n.d.). However, while a valid measure of critical speed, the reliability of the Maryn 3-minute test remains unknown.

CHAPTER 2 – LITERATURE REVIEW

Purpose and Scope

The purpose of this literature review is to provide background knowledge relevant to the study that was performed. This review addresses: 1) the importance of measuring fitness, 2) different measures of cardiovascular fitness, 3) critical speed, 4) critical speed fitness tests, 5) the 3-minute all-out critical speed test, 6) the application of critical speed for prescribing exercise intensities, and 7) the gap in the knowledge.

Importance of Measuring Physical Fitness

Physical fitness can be thought of as an integrated measure of nearly all of the body's functions utilized throughout the execution of daily activities and/or physical exercise and therefore is considered as an important health marker both in the early years and later in life (Ortega et al., 2008). Being physically active has many health benefits and a lack of physical activity is a modifiable risk factor for cardiovascular disease and a wide variety of other chronic diseases, including but not limited to obesity, hypertension, and depression (Warburton et al., 2006).

In the domain of physical fitness there are five separate components: muscular endurance, muscular strength, cardiovascular fitness, flexibility, and body composition (Caspersen & Christenson, 1985). Each of these components has specific measures that accurately represent them along with varying methods to achieve those measures. Having the measures and methods to represent a person's level of fitness is important for many reasons. These reasons are mostly applicable to athletes and training populations that are interested in enhancing performance and improving from their baseline measures, whether for health or sport purposes. For individuals interested in performance, these measures provide a starting point that allows them to set reasonable goals and to track their progress. Such measures can also aid in selecting appropriate work and training loads, and in some cases these measures can also assist in the detection of potential injury (June et al., 2011). The focus of this review is on cardiovascular fitness which can be understood as "the ability of the circulatory and respiratory systems to supply fuel during sustained physical activity and to eliminate fatigue products after supplying fuel" (Caspersen & Christenson, 1985).

Different Measures of Cardiovascular Fitness

In measuring a person's cardiovascular endurance fitness, the parameters most commonly used include VO_2 max, exercise economy, and blood lactate threshold (Jones & Carter, 2000). VO_2 max represents a person's highest rate of oxygen consumption attainable during maximal exercise. VO_2 max is widely regarded as the gold standard of measuring a person's maximal oxygen uptake. It refers to the maximum capacity to transport and utilize oxygen during exercise done at increasing intensity. While VO_2 max is regarded as the gold standard for aerobic fitness measurements, the translation of the VO_2 max measure to real-life application can be complicated given that VO_2 max is expressed in units of liters or milliliters per minute (Shete, 2014) and that is not how athletes or fitness professionals typically describe exercise intensities.

Exercise economy is known as "the oxygen uptake required at a given absolute exercise intensity" (Jones & Carter, 2000). An endurance athlete with good economy uses less oxygen than an endurance athlete with poor economy at the same steady-state speed (Thomas et al., 1999). Exercise economy is really a complex and multifactorial concept that incorporates metabolic, biomechanical, neuromuscular, and cardiorespiratory factors (Barnes & Kilding, 2015). Lab testing for exercise economy is performed on a treadmill while gas-exchange is determined. However the results on the treadmill are typically under-estimations of the true energy demands required and field testing is impractical due to the influence of changes in environmental conditions (Barnes & Kilding, 2015). With so many factors to consider, the testing requirements for exercise economy create quite the hurdle for real-world application.

Blood lactate concentration is another parameter often used to measure a person's cardiovascular endurance. In testing, blood lactate concentration measurements are acquired while exercise work rate increases. At first, the blood lactate concentration increases gradually, but as the exercise becomes more intense the concentration rises more rapidly. The training intensities associated with the increase in blood lactate above resting levels is an accurate predictor of endurance performance (Jones & Carter, 2000). However, blood lactate concentrations require invasive methods that involve drawing blood from the participant, which can be a turn-off for some people.

Critical speed is another indicator of a person's aerobic fitness. It is expressed in units of meters per second and can potentially only require a single-visit running test (Di Prampero, 1999). What separates critical speed from the previously mentioned measures of cardiovascular endurance is how well it translates to training and exercise and how it can be a more accessible measure to acquire through field testing (Weir, 1997).

Critical Speed

In regards to cardiovascular exercise, there are three intensity domains: moderate, heavy, and severe (Gaesser & Poole, 1996). The boundary between the heavy and severe domains is marked by critical speed. Critical speed is defined as “the tolerable duration of severe intensity exercise”, and “the greatest metabolic rate that results in the rate of lactate production being matched by the rate of clearance” (Poole et al., 2016). Critical speed is one of two parameters that make up the hyperbolic relationship between speed and duration. There is a curvilinear relationship of time to the limit of tolerance plotted against constant speed outputs (Poole et al., 2016).

Critical speed represents the highest sustainable work rate, and the curvature constant (D’), represents the maximum distance that can be covered above critical speed (Vanhatalo et al., 2007). The speed-duration relationship that critical speed represents has many uses in helping to understand fatigue and mechanisms of fatigue across the lifespan in healthy individuals and patient populations. It takes into account the effects of environmental challenges on exercise tolerance, and evaluates the effectiveness of therapeutic countermeasures (Poole et al., 2016). It also precisely predicts exercise performance, determines the feasibility of an athletic feat, and can model optimal performance tactics for a team, or group of athletes, that has different critical speed or D’ values (Poole et al., 2016).

Critical Speed Fitness Tests

Traditionally, critical speed and D’ were determined through a series of different runs to exhaustion at various intensities. The participants would first go through a VO_2 max test to determine the speed that elicits their VO_2 max. Following that test, the participants would complete runs to exhaustion at speeds of 90%, 100%, 120%, and 140% of the speed that elicited their VO_2 max (Di Prampero, 1999). Other test models to determine critical speed have been seen in cycling ergometers and swimming. In the cycle ergometer model, participants performed constant-intensity tests to exhaustion on the ergometer. They were instructed to perform as much work as possible during trials that lasted 4, 9, and 14 minutes with initial work rates ranging from 85-95% of their maximum power (Dekerle & Paterson, 2016). In the swimming model, participants were required to complete three timed maximal-effort swims at distances of 200, 400, and 1000 meters. The three tests were completed on successive days (Barden & Kell, 2009).

Reliability studies have been conducted on these traditional models of determining a person's critical speed. A study conducted by Galbraith et al. (2014), found that a test consisting of multiple maximal runs to exhaustion to determine critical speed was reliable across repeated tests. A similar study confirmed the reliability of critical speed testing for swimming models that required three different races over distances of 200-, 400-, and 800-meters to determine critical speed (Dekerle & Paterson, 2016).

While a reliable and valid method of determining a person's critical speed, these types of tests are time consuming, have been found to be problematic for certain study designs, and as a result an all-out exercise test that can be completed in a single testing session was developed (Burnley et al., 2006). This all-out exercise test for determining critical speed had a duration of 3 minutes as this is long enough to yield a stable power output at the end of the test, but not so long that subjects fail to complete the test (Burnley et al., 2006).

Vanhatalo et al. (2007), found that the single 3-minute test performed on a cycle ergometer could produce accurate critical power and W' values (Vanhatalo et al., 2007). Pettitt et al. (2012) confirmed the validity of conducting a running 3-minute test through GPS data compared to the graded exercise test with verification for VO_2 max (Pettitt et al., 2012), and Broxterman et al. (2013) confirmed the validity of determining critical speed using the 3-minute running test with use of an accelerometer compared to the traditional incremental test method (Broxterman et al., 2013). Then, in 2013, Clark et al. assessed the use of critical speed to prescribe appropriate training loads to female soccer players and conducted the 3-minute test with the use of a camera recording system to achieve accurate critical speed and D' measures (Clark, 2013).

3-Minute All-Out Critical Speed Test

For the purpose of this study, the focus will be on application of the 3-minute all-out running test for determining critical speed. Participants perform the 3-minute running test on a track, starting in a stationary standing position. At the start, participants immediately work into their top sprint speed and seek to maintain that level of speed throughout the duration of the 3 minutes. Eventually, participants will see decreases in their speed, but they continue to run as hard and fast as they can. While the procedures for completing the test are generally consistent, previous studies utilizing the 3-minute test for running have differed with respect to the measurement tools used (GPS, camera system, accelerometer) to quantify time and/or distance.

Pettitt et al. (2012) conducted their test on a level outdoor 400-m track with minimal wind conditions and a clear sky. The subjects wore a GPS wrist watch, which was used to track their displacements covered at 150 and 180 seconds (Pettitt et al., 2012). Broxterman et al. (2013) performed their test on an outdoor 400-m track, and recorded their subjects' speed using an accelerometer placed on their right foot (Broxterman et al., 2013). In both tests, subjects were provided with strong verbal encouragement throughout and were unaware of elapsed time nor time remaining to prevent pacing. Despite using different measurement tools, the investigations conducted by Pettitt et al. (2012) and Broxterman et al. (2013) found no significant differences in the attained critical speed and D' values and the values obtained using the traditional method - multiple graded exercise treadmill runs to exhaustion.

These 3-minute tests help to address the issue of practicality in regards to time. As the 3-minute test is conducted during a single session, it eliminates the need for multiple test bouts across numerous occasions. However, there are still some challenges present preventing it from becoming more widely accepted and more suitable for the populations that would be most likely to use it, such as coaches, personal trainers, and athletes. While GPS, accelerometers, and cameras provide detailed and accurate position data, these materials are costly and therefore limit the accessibility of the test. However, the calculation of critical speed and D' is only dependent upon two variables collected from testing: distance and time. As long as distance and time are tracked, critical speed and D' can be determined. Accordingly, Maryn et al. recently evaluated the validity of conducting the test using just cones and a stopwatch instead of expensive equipment. By placing cones at 50 meters intervals around a 400-meter track, it is possible for a tester to conduct the 3-minute all-out test using just a stopwatch to record the split times at each passed cone covered by the subject from the start to the end of the test. Conducting the 3-minute test in this manner is feasible with respect to access to resources. Maryn et al., confirmed the validity of the 3-minute test using cones and a stopwatch by comparing their results to laboratory test measures of critical speed (Maryn et al., n.d.). However, they did not assess the consistency of this test across days and testers. Therefore, the purpose of this study is to examine the test-retest and inter-rater reliability of the 3-minute test as conducted by Maryn et al.

Application

The availability of a critical speed fitness test that is practical with respect to both time and required resources could lead to greater widespread application of the test. If we are able to confirm that critical speed determined using the Maryn 3-minute test is consistent across different testing sessions, it will indicate that any changes in critical speed identified in people are likely due to changes in cardiovascular fitness and not due to practice effects from completing the test multiple times. The identification of high inter-rater reliability will indicate that the test can be conducted by anyone with minimal training, and that it does not require someone with extensive training in order to obtain a valid result. If we are successfully able to demonstrate that the Maryn 3-minute test can produce reliable estimates of critical speed independent of testing day and rater then that would eliminate one of the major hurdles preventing critical speed testing from becoming a more common form of cardiovascular fitness testing. Moreover, one of the advantages of critical speed and more specifically D’ is that the test result can be used to directly inform exercise prescription.

One example of critical speed and D’ being utilized to inform exercise prescription was researched by Clark et al. (2013) when he examined the effectiveness of using the critical speed derived from the 3-minute test to prescribe high-intensity interval training amongst female college soccer players. Based on the subjects’ initial critical speed and D’ values, they were assigned to either a low critical speed or a low D’ group relative to the teams mean speed velocity and D’. The four-week training programs prescribed were designed to make the team more homogenous for both critical speed and D’. At the conclusion of the four-week training protocol, both groups remained significantly different from each other at pretesting and post-testing for critical speed and for D’, but the team as a whole experienced an increase in critical velocity by 0.22 m/s (Clark et al., 2013). This is just one of the experimentally tested ways that the 3-minute all-out test can provide critical speed measures that can be used for exercise prescription.

Gap in Knowledge

It is evident that the determination of critical speed is a useful indicator of cardiovascular fitness that can be used to directly inform exercise prescription. However, while Maryn et al. has demonstrated the validity of a practical field test for determining critical speed, it is unknown if this test is reliable across days and across raters. If the Maryn 3-minute running test were shown to be reliable, then the results would support the use of this test to determine critical speed in a time efficient manner and without the need for expensive test materials. This test, which only requires: 3 minutes of testing time, a stopwatch, cones, and one individual to administer the test is very practical for coaches, fitness professionals, and even athletes themselves to use. Additionally, critical speed provides a better “understanding of the limitations to human performance and the fatigue processes that underpin them” and with the use of a person’s critical speed, practitioners will be able to develop effective training programs to optimize performance and minimize risk of injury (Poole et al., 2016). Accordingly, the 3-minute all-out test could become a more popular and common choice of fitness test to measure a person’s cardiovascular fitness. This has the potential to revolutionize the way that cardiovascular fitness testing is conducted saving teams and businesses time and money. At the same time, it could also provide them with a parameter that reveals more to them about their level of fitness and is easier to translate to their training loads and prescriptions. In the end, we could expect to see more athletes that are healthier and better trained consequently leading to better performances.

CHAPTER 3 – METHODS

Participants

Data was collected from 16 participants ranging in age from 18-35 years old. An *a priori* power analysis indicated a minimal sample size of n=9 would be required to detect an ICC of 0.7 with 80% statistical power with a significance level of alpha ≤ 0.05 (Adam & Baharum, 2017). Participants were volunteers from nearby regions. The study performed by Maryn et al., found that the 3-minute test provided valid measures of critical speed in healthy, recreationally active individuals (Maryn et al., n.d.). As a result, we only recruited individuals that reported regularly participating in cardiovascular training a minimum of three times a week and that they could complete a run of at least three miles at a pace faster than 10 minutes per mile. Ideally, this meant that our participants were at the very least recreationally fit and active, and that they would be similar to the individuals that produced valid measures of critical speed in the study conducted by Maryn et al. Participants who reported currently having any injury or illness that restricted their physical activity or that any medical provider had recommended restrictions on their physical activity that would include not running as fast as possible for three minutes were excluded from participation. We also screened potential participants for risk factors for cardiovascular disease using exclusion criteria that were drawn from the American College of Sports Medicine's "Risk Stratification Screening Questionnaire" (ACSM, n.d.). All participants provided written informed consent and were asked to avoid strenuous exercise that was outside of their normal training for 24 hours before each test session.

Experimental Design

A test-retest design was implemented to establish reliability across days. In addition to completing the 3-minute all-out test themselves, participants in this study served as raters as other participants performed the test. The results obtained by these raters were compared with the results obtained by the principal investigator (L.A.) to determine whether the CS could be accurately assessed by raters with a limited amount of training.

Participants completed an eligibility screening and familiarization session and two testing sessions that were separated by a minimum of 3 days (72 hours) and a maximum of 14 days. This was done to ensure that the participant was able to perform to the best of their ability. It has been shown that delayed onset muscle soreness can negatively affect athletic performance (Cheung et al., 2003), and therefore, a minimum time of 72 hours was selected between test sessions as that is when delayed onset muscle soreness is reported to subside (Armstrong, 1984). The maximum time between sessions of 14 days was selected as a range of time that allowed for convenience in scheduling a second session with the participant as well as being a short enough period of time to minimize the potential that any between-day differences in critical speed that may be observed would be due to changes in maximal aerobic variables associated with a training effect (Pollock et al., 2015) rather than measurement error associated with the Maryn 3-minute test. During the days in-between test sessions, participants were encouraged to maintain their regular physical activities to avoid a detraining effect.

In addition, due to the fact that athletic performance can be affected by time of day (Teo et al., 2011) and environmental conditions, each participant performed both of their test sessions in the same period of the day and under the following environmental conditions. For temperature, several studies have shown that aerobic performance decreases as ambient heat nears temperatures of ≥ 30°C (Tyler, 2008);(Morris et al., 2017);(Altareki et al., 2009). Therefore, we did not conduct tests in temperatures over 25°C (77°F) to avoid our participants' performance being negatively affected by the heat. Similarly, performance can be negatively affected in severely cold weather where a person is unable to keep their core temperature at 35°C (Faulkner et al., 1981);(Castellani et al., 2012). Data from the Winter Olympic Games reveal that a person's core temperature can be maintained at levels safe from hypothermia in 4°C weather as long as they sustain a working intensity of at least 60% VO_2 max for a duration around 3 minutes (Castellani et al., 2012). Therefore, we did not conduct tests in temperatures below 4°C (39°F). Moreover, a study that analyzed the effects of temperature on marathon running performance found that every increase of 1°C from an optimal temperature of 9.9°C will result in a speed loss of 0.03% (Helou et al., 2012). Therefore, within our set testing range of 4-25°C the maximum change we could expect to see in an individual's performance due to temperature would be 0.63%.

For wind, we took into consideration the fact that according to the International Association of Athletics Federations rules, if the average wind velocity measures to be greater than 2 m/s, performance results could be considered invalid if it is determined that the results were affected unfairly due to wind assistance (IAAF, 2017). However, we also wanted to maintain the practicality of being able to conduct this as a field test and felt that it was unrealistic to only be able to test when wind speeds were less than 2 m/s, which translates to 4.47 mph. According to the Beaufort wind scale, wind speeds between 13-18 mph are considered a moderate breeze and wind speeds up to 24 mph are considered light winds (Beaufort, 1805). As a compromise, we chose to conduct tests only when the sustained wind at the testing site was between 0-10 mph. Finally, testing was conducted regardless of whether there was precipitation, so long as there was no standing water on the track.

Procedures

Participants completed a single eligibility and familiarization session (Session 1) and two testing sessions (Sessions 2 and 3) on the Oregon State University campus.

<u>Session 1: Eligibility and familiarization</u>

All study participants completed an eligibility screening in the Women's Building on the Oregon State University campus after informed consent was obtained. First, the individual completed the first two pages of the participant screening questionnaire (see **Appendix One**) by themselves. A research team member then reviewed their answers to confirm that he/she met the general criteria for enrollment as defined by being 18-35 years of age; answering YES to questions 3 and 5; and NO to questions 6-16. Participants that met all of these criteria then had their height and weight measured and their body mass index (BMI) calculated. Participants with a BMI < 30.0 kg/m^2 were eligible for continued participation and immediately completed a familiarization session. No participants were determined to be ineligible by the screening process.

After eligibility had been confirmed, a member of the research team described to the participant that the 3-minute running tests would be conducted on a 400-m track with cones placed 50 m apart (**Figure 1**). They were told that the role of the rater for this test is to record the split time as the participant passes by each cone for the duration of the test. They were familiarized with the data collection sheet and where they would write down the split times when serving as the rater. Participants were also familiarized with the stopwatch that they would use at the testing sessions to measure split times and were provided with time to practice using the stopwatch.

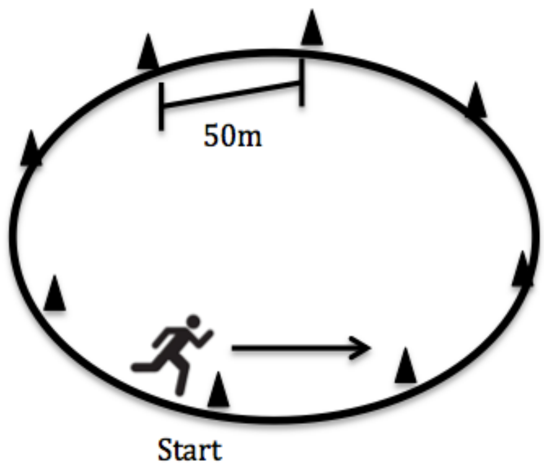

Figure 1: Test outline of the set-up for the Maryn 3-minute test. Participant is ready to begin the test at the starting point, running around the 400m track with cones placed every 50m.

Sessions 2 and 3: Running test sessions

Prior to arriving at the testing site, the research team evaluated temperature and wind speed using the online national weather service report AccuWeather: *www.accuweather.com*. This information was used to make an initial determination as to whether or not the environmental conditions would meet the previously described requirements.

After confirming that the weather conditions were likely appropriate for testing, the research team arrived at the testing site. Testing was performed on an outdoor, 400-meter synthetic track (Whyte Track and Field Center at Oregon State University) to replicate the surface used by Maryn et al. In preparation for testing, the principal investigator would confirm that the weather conditions at the site met the requirements using an anemometer (Benetech, Palo Alto, CA) and placed cones around the track at every 50 meters. Starting at the 0 meter point, using a measuring wheel (Keson, Aurora, IL), the test conductor placed one orange cone down every 50 meters, on the infield just inside of Lane 1, until they returned to the starting point.

On the day of each testing session, participants met with the research team at the testing site. Participants completed one trial of the 3-minute all-out running test and served as a test rater for one other participant during each testing session. The actual wind speed (mph) and temperature (°F) prior to each trial was measured using the anemometer and recorded by the principal investigator.

3-minute Running Test: Prior to completing the all-out running test, participants were taken through a standard 15-minute dynamic warm-up that was intended to prepare the participants to perform the 3-minute test, but should not have elicited any fatigue (**Table 1**) (Maryn et al., n.d.).

Table 1. Dynamic Warm-up Protocol

Light jog	Once around track
High Knees Butt Kicks Side Shuffles (both sides) Karaokes (both sides) Back Pedal	1 x 20 yards each

Knee Pulls to Chest Quad Stretch Toe-Touches Lunge with Twist Side Lunges Arm Scissors Arm Circles Leg Swings	 10 times each side

To complete the running test, participants were instructed to start in a stationary upright standing position on the inside lane of the track at the starting mark. Once instructed to begin running, they were to quickly work up to running at their maximal sprint speed and continue to run at maximal effort in the inside lane of the track for the duration of the test (3 minutes). Once these instructions were made clear, participants then performed a single 3-minute test. Strong verbal encouragement was provided throughout the test by a research team member, although participants were not made aware of the elapsed time nor time remaining in an effort to prevent pacing. During the test, both a research team member and one other participant recorded the runner's performance as described below. At the completion of the test, participants were instructed to walk around the outside lane of the track and monitored until they recovered from their effort.

Test Rating: When serving as a rater for a running test trial, participants positioned themselves along with a research team member on the middle of the field at the center of the track to best view the runner completing the test. Test raters for each trial measured the total time and time splits at each 50 m interval using a stopwatch and recorded the information on the data collection form **(Appendix Two).**

Data Reduction

Using the average of the time splits recorded over the last two cones passed in the 3-minute test, we calculated the participant's critical speed using the following formula: CS=50m/time (Maryn et al., n.d.). As an example, if the last two recorded time splits for a participant were 13 seconds and 14 seconds, respectively, the calculated average is 13.5 seconds. Critical speed in this example would equal 50 meters divided by 13.5 seconds, or 3.70 m/s.

Statistics

Test-retest reliability (Specific Aim 1) was determined using an ICC model (2,1). The strength of agreement for the ICC ranges was interpreted as follows: 0.00 to 0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial; and 0.81 to 1.00, almost perfect (Landis & Koch, 2008). The critical speed test measures as determined by the research team member for the 11 participants that completed the two test sessions on different days were included in the analysis for Aim 1.

Inter-rater reliability (Specific Aim 2) was assessed using an ICC model (1,1) to account for random rater effects (i.e., potential for different pairs of raters for each observation) using the critical speed values determined from only the first testing session of all 16 participants. The strength of inter-rater agreement as assessed by the ICC model was interpreted using the same ranges.

Finally, to examine the differences in CS between test sessions (Aim 1) or raters of the same test (Aim 2), we generated Bland-Altman plots to show systematic and/or random error differences for test-retest reliability between days and for inter-rater reliability during testing session 1 (Bland & Altman, 2010). In addition, 1-sample t-tests were used to determine whether the differences in CS between days and between raters, respectively, was significantly different than zero. All data was analyzed using SPSS (Version 23; Chicago, IL) and Microsoft Excel (Version 2016; Redmond, WA).

CHAPTER 4 - RESULTS

The descriptive characteristics of the participants are presented in **Table 2**.

Table 2. Mean descriptive characteristics of participants.

n	Age (y)	Height (m)	Mass (kg)	BMI
16	25.75 ± 4.89	1.73 ± 0.10	70.58 ± 15.03	23.4 ± 3.34

n = 13 males and 3 females, BMI = body mass index.

The environmental conditions recorded, differences between sessions, and ranges for the wind speed, temperature, and time are presented in **Table 3**. There was no precipitation during any of the testing sessions.

Table 3. Test-retest controlled conditions.

Wind Speed (mph)			Temp (°F)			Time Between (hrs)	
AVG±SD	AVG DIFF	RANGE	AVG±SD	AVG DIFF	RANGE	AVG±SD	RANGE
2.6±1.5	1.38	0.7-6.7	68.4±5.2	5.45	59-76	93.8±19.7	72-120

AVG±SD = average from all test sessions plus/minus the standard deviation, AVG DIFF = average of the differences between test sessions 1 and 2, RANGE = minimum and maximum values.

Critical speed measures recorded from all test sessions are presented in **Table 4**.

Table 4. Maryn 3-minute test results.

Subject	Test 1 CS (m/s)		Test 2 CS (m/s)	INTER-RATER DIFF (m/s)	BETWEEN-DAY DIFF (m/s)
	ER	PR	ER		
1	3.16	3.14	3.68	0.024	-0.521
2	3.92	3.90	4.13	0.017	-0.212
3	4.02	4.02	4.51	0.002	-0.487
4	3.43	3.38		0.048	
5	3.00	3.02	3.84	-0.017	-0.839
6	4.10	4.09	4.58	0.013	-0.480
7	3.35	3.36	3.49	-0.010	-0.143
8	3.36	3.36	3.85	0.007	-0.490
9	5.26	5.25	5.32	0.017	-0.062
10	3.38	3.39	3.87	-0.010	-0.486
11	4.84	4.94		-0.094	
12	3.41	3.39		0.019	
13	4.88	4.87	5.26	0.014	-0.378
14	3.03	3.07	3.47	-0.045	-0.437
15	2.73	2.73		0.005	
16	4.03	4.03		-0.003	

CS = critical speed, ER = expert rater, PR = participant rater, INTER-RATER DIFF = difference between ER and PR for Test 1, BETWEEN-DAY DIFF = difference between ER Test 1 and Test 2.

For Aim 1, which assessed the test-retest reliability of the test, the ICC (2,1) = 0.82 with a standard error of the measure (SEM) of 0.32 m/s. For Aim 2, which evaluated the inter-rater reliability of the test, the ICC (1,1) = 0.99 with a SEM = 0.02 m/s.

Presented below are Bland-Altman plots for test-retest reliability (**Figure 2**) and inter-rater reliability (**Figure 3**). As indicated by these plots, we identified a significant difference in the average CS value between sessions (CS difference = 0.042 ± 0.21 m/s, $p <$ 0.001), but not between raters (CS difference = 0.001 ± 0.03 m/s, $p = 0.909$).

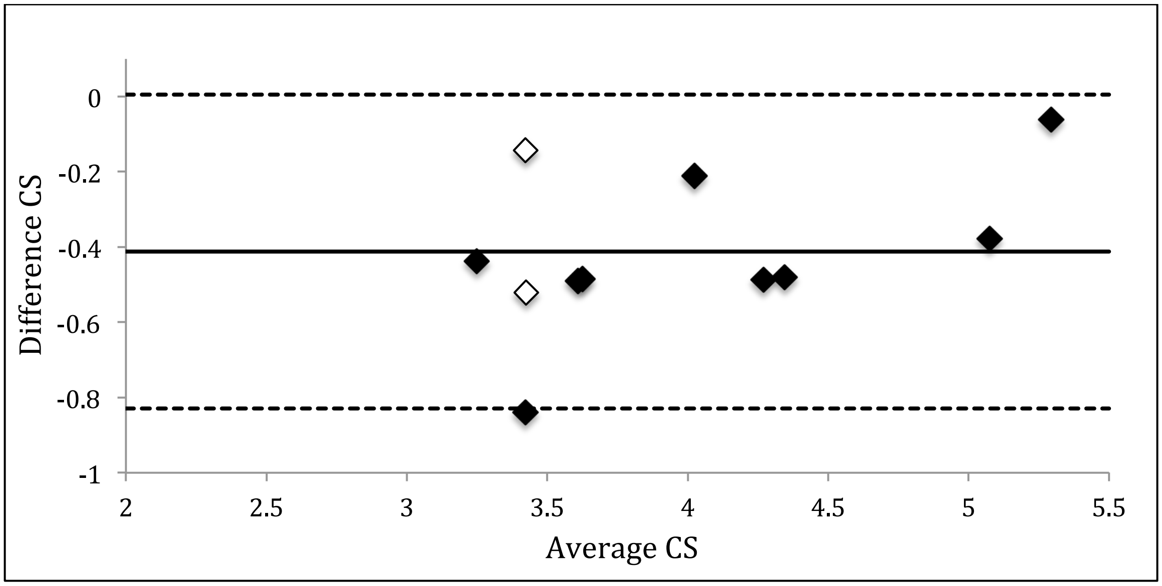

Figure 2. Bland-Altman plot for test-retest reliability (Aim 1) with the difference of the critical speed (CS) scores from test session 1 and test session 2 recorded by the expert rater plotted against the average of the same scores. White markers represent female subjects while black markers represent male subjects. The solid, center line demonstrates the average of the differences from the two test sessions. The outer dashed lines mark the 95% limits of agreement. All units are in m/s.

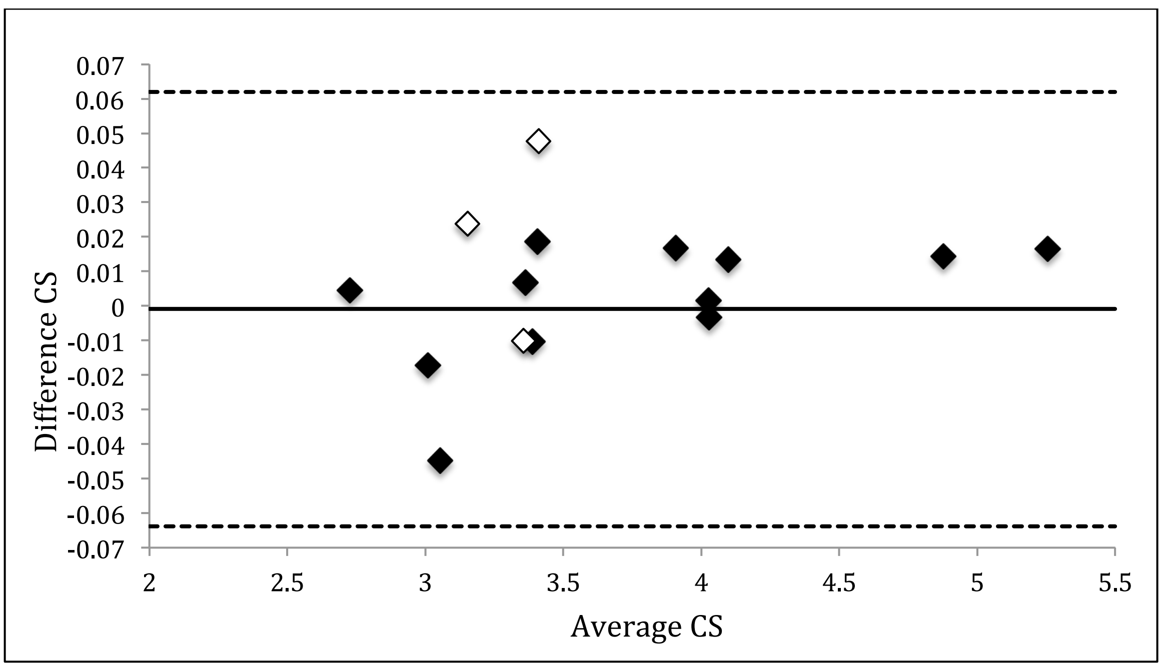

Figure 3. Bland-Altman plot for inter-rater reliability (Aim 2) with the difference of the critical speed (CS) scores recorded between the expert rater and the participant rater from test session 1 plotted against the average of the same scores. White markers represent female subjects while black markers represent male subjects. The solid, center line demonstrates the average of the differences from the two test sessions. The outer dashed lines mark the 95% limits of agreement. All units are in m/s.

CHAPTER 5 - DISCUSSION

The primary findings of this investigation are that the use of the Maryn 3-minute test as applied in this study requires no specialized training as evidenced by the fact that the CS values obtained by individuals with minimal training were consistent to those obtained by an expert rater. However, while the test-retest reliability of the 3-minute test was strong-excellent (ICC = 0.82), there was a systematic bias detected in which participants exhibited significantly greater CS values during testing session 2 compared to testing session 1, and the magnitude of this difference (0.41 m/s) is likely clinically meaningful.

With respect to the need for trained personnel to administer the test, our results indicate that the CS values obtained by an expert and untrained rater using the Maryn 3-minute test exhibited almost perfect consistency (ICC= 0.99) and the mean difference in CS between raters was negligible (< 0.001 m/s, Figure 3). The critical speed measures recorded by the expert and participant raters in this study (3.75 m/s) were also comparable to the average values of 3.82 m/s (Maryn et al., n.d.), 4.06 m/s (Galbraith et al., 2014), and 4.46 m/s (Pettitt et al., 2012) reported in previous studies that assessed critical speed during running. Overall, the inter-rater reliability results suggest that this test could be a valuable option for people such as coaches, personal trainers, etc. to assess CS. The testing equipment required for this study was very easily acquired and the set-up for each test took only a few minutes, which is in contrast to the need for the high-cost equipment (Pettitt et al., 2012);(Broxterman et al., 2013) or multiple days of testing (Di Prampero, 1999) of other CS testing protocols. Further, the participant raters in this study received no formal training on collecting data other than being familiarized with the stopwatch, data collection sheet, and testing procedures. Despite this, the critical speed measures of the expert rater and participant raters were highly consistent for the same trial. This shows that the test does not require a specially-trained rater to conduct the test. For practical application, this can be especially useful for sport teams. Instead of relying on a single, highly trained rater who would need to test each player one-by-one, it is possible coaches and/or strength and conditioning professionals could pair players up and these players could act as raters for each other. Being able to do so could make the application of this test feasible by saving lots of time, and in so doing this test could be an attractive fitness test option for teams with lots of athletes and/or limited resources.

However, while the Maryn 3-minute test can be easily administered and rated by untrained individuals, the results related to the test-retest reliability do not support the use of this test for determining CS when the test is administered as described in this study. Though the test-retest reliability of the test was strong-excellent (ICC = 0.82) and greater than the 0.7 threshold deemed prior to the study as clinically meaningful, we observed a systematic bias whereby participants exhibited an average CS that was 0.41 m/s faster during the second testing session compared to the first testing session (Figure 2). To look at these results from a more practical perspective, the average critical speed recorded by the expert rater from the first test session translates to a 7-minute and 6-second mile pace while the average critical speed recorded by the expert rater from the second test session translates to a 6-minute and 25-second mile pace. This suggests that the magnitude of this difference in CS between testing sessions is likely meaningful when making inferences about cardiovascular fitness with this conclusion supported by previous reports.

Nimmerichter et al. (2015) compared CS between trained and untrained participants classified as having a regular training volume of at least 6 hrs/week with the training including participation in a variety of team and individual sports or a regular training volume less than 3 hrs/week, respectively. They found that the average difference in critical speed between the untrained and trained participants was 0.673 ± 0.107 m/s (Nimmerichter et al., 2015). Though this study is limited by the fact that fitness status was determined by self-report, Clark et al. (2013) administered high-intensity interval training programs to a team of female soccer players twice a week for four weeks and examined the change in critical speed. They found that the team average at the start of the study was 3.68 ± 0.24 m/s and after the four weeks the team average increased to 3.82 ± 0.21. Therefore, after a four-week training program, the group increased their critical speed by just 0.22 m/s on average (Clark et al., 2013). Finally, Kramer et al. (2018) used a large sample of normative data from rugby players to generate categorizations to rank critical speed values from extremely low to extremely high with the average difference between ranks being 0.28 ± 0.17 m/s which suggests that the Maryn 3-minute tests as administered in this study is not able to detect small changes in cardiovascular fitness levels. However, though the average difference between the smaller categorizations was 0.28 m/s, the magnitude of the difference between the broader “low”, “average”, and “high” categories was larger (0.44 m/s). Therefore, while the magnitude of the difference we detected between-days may be too large to accurately categorize or detect changes between detailed categories (e.g., from “below average” to “average”) the test may be able to discriminate between broader categories such as “average” and “high” (Kramer et al., 2018). While this may be useful for some applications, the fact remains that our participants should not have had any change in their CV fitness between testing sessions and the relatively large magnitude of the difference in CS value between sessions could incorrectly be interpreted as small, but potentially meaningful changes in fitness rather than due to other factors unrelated to fitness level. Although it is not possible to identify exactly what factor(s) may have driven the improved scores in testing Session 2, there are some potential factors like pacing, motivation, and a test practice effect that should be considered.

The participants could have paced themselves during the second session rather than fully exerting themselves throughout the entire duration of the test, which could have led to better critical speed scores in the second test sessions. The 3-minute test only produces an accurate measure of an individual's critical speed if they follow the procedures correctly by working up to their sprint speed as quickly as they can from the start and continuing to try to maintain maximal speed for the whole three minutes. By running all-out, the participant becomes fully depleted as they near the end of the test and are left running at their critical speed. However, if a participant does not fully deplete themselves during the initial portions of the test, then it is possible for them to finish the test with faster split times resulting in a faster calculated critical speed score. For practitioners to be able to use this test to accurately assess CS, there needs to be a way to prevent or detect pacing without making the test too difficult or complicated for widespread use. While we chose to evaluate the test using minimal controls (i.e., to not include a specific check for pacing in order to mirror the simplest test administration possible), it might be possible to teach coaches how to detect pacing using the data as it was recorded in this study. There should be a continuous increase in the time needed to complete each split. If an individual's split times towards the end of the test are shorter than they were in the middle of the test, then this is an indication that they were pacing.

However, this would not address a situation where a person could exhibit split times that increase throughout the test, but not be running all-out, which would also result in an inaccurate critical speed measure. Therefore, a better solution to identifying pacing may be to determine a percentage threshold of how much of a drop-off a person should expect to experience from their first couple of splits to their last. Though it is highly likely that there needs to be stricter checks built-in to control for and identify pacing, there also needs to be a balance maintained so that the administration of the test does not become too complicated and challenging to be practical.

It is also possible that participants experienced a motivation effect. Knowing that they would run the same exact test twice within a short span of time, it is possible that the participants felt intrinsically motivated to outperform themselves on their second test trial. Several participants expressed a sense of accomplishment when they were able to finish further around the track compared to their first running trial. This desire to improve on their own previous performance could have possibly led to greater effort and higher critical speed scores. Similarly, a test practice effect could have contributed to the faster critical speed scores on the second test session compared to the first test session. Even though the test was designed to have the participant be unaware of how much time has passed or how much time was remaining during their test, it is possible that they were able to use their experience from the first test to help them on the second test. If a participant remembered that they ended at 750 meters on their first test, as they reached 650 m or 700 m during their second test they would know that they were close to the end of the three minutes.

Using this knowledge, the participants could speed up a little bit more towards the end knowing they are almost done which would reduce their final split times and could possibly explain the increase in critical speed measures for participants on their second test session.

One possible solution to the potential motivation and test practice effects could be to have the participant begin the test at a different 50m mark on the track for the second test session. This could also throw off any possible connection for the participant between positional-awareness and elapsed time. Future studies should also evaluate if any motivation and/or practice effects only influence CS scores between the first and second administrations of the test by evaluating if there is consistency in the critical speed measures between CS scores obtained between the second and third administration of the test.

As with all investigations, the current study was not without limitations. All participants received the same verbal encouragement for each of their test trials. The verbal encouragement came from the expert rater on both test sessions 1 and 2 and from the expert rater and participant rater on test session 1. The raters shouted words of encouragement at the direction of the runner from the center of the field inside the track. However, it is possible that the verbal encouragement was easier to hear on stretches of the track where the runner was in closer proximity to the raters. It is also possible that participants who were intrinsically motivated to perform well responded to the verbal encouragement differently then those participants who were not motivated. If the participant had no intrinsic desire to perform well on the test, then it is unlikely that the added words of encouragement would result in a greater effort.

In addition, we utilized a relatively broad set of inclusion criteria that resulted in a sample that was slightly different than those of previous investigations. The mean age of our participants was 25.75 ± 4.89 years with an average height of 1.73 ± 0.10 m and mass of 70.58 ± 15.03 kg. Of our 16 participants, 13 were male and 3 were female. While a few of our participants were competitive runners, the majority were recreationally active. Similar studies had subject populations that were as follows: 20 NCAA Division II Women's Soccer players with an average age of 19 ± 1.0 years, height of 1.68 ± 0.06 m, and mass of 61 ± 6.0 kg (Clark et al., 2013), 14 collegiate women distance runners with an average age of 19 ± 1.0 years (Pettitt et al., 2012), 7 subjects (4 males and 3 females) with an average age of 25.3 ± 3.4 years, height of 1.74 ± 0.11 m, and mass of 69.7 ± 13.7 kg that ranged from active to highly trained (Broxterman et al., 2013), and 24 subjects (11 males and 14 females) of all fitness levels with an average age of 22.9 ± 4.65 years, height of 1.75 ± 0.08 m, and mass of 72.45 ± 12.80 kg (Maryn et al., n.d.). Given these discrepancies in participant characteristics, it is possible that the between-day reliability and magnitude of variability between testing sessions identified in this study may not be the same when the test is used with participants of different ages, fitness status, and/or sport participation.

CHAPTER 6 – CONCLUSION

As an outcome of this study, we expected to find, in aerobically trained college-aged men and women, consistent critical speed measures from the Maryn 3-minute test across different days and different raters. Though we found the Maryn 3-minute test to have high reliability between raters, the critical speed measures obtained between days were significantly different and the magnitude of this difference was large enough that it may lead to different interpretations of an individual's cardiovascular fitness. However, future research is needed to determine if the 0.41 m/s difference is meaningful with respect to differentiating fitness levels of individuals. Therefore, while the test can be conducted consistently by anyone with minimal familiarization with a stopwatch and the data collection methods, the use of this test for determining critical speed by coaches, physical education teachers, or personal trainers without the addition of greater testing controls is not supported. Future research should attempt to address potential issues related to pacing by identifying the minimal amount of added controls needed to produce consistently accurate measures of critical speed.

If future studies succeed in finding that optimal balance in pace-restricting controls and practicality for widespread use, it might facilitate an increase in the use of critical speed as an indicator of aerobic fitness by coaches and fitness professionals due to the convenience of the test being short in time and low in cost. An increased utilization of critical speed in athletic settings can help athletes that are concerned about their aerobic fitness by leading them to more optimized training programs and better athletic performance.

BIBLIOGRAPHY

ACSM. (n.d.). ACSM Risk Stratification Screening Questionnaire. Retrieved from http://www.exerciseismedicine.org/assets/page_documents/Appendix D - ACSM Risk Stratification Q.pdf

Adam, M., & Baharum, N. (2017). A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient : a review, *12*, 1–11.

Altareki, N., Drust, B., Atkinson, G., Cable, T., & Gregson, W. (2009). Effects of Environmental Heat Stress (35 8 C) with Simulated Air Movement on the Thermoregulatory Responses during a 4-km Cycling Time Trial. *International Journal of Sports Medicine*, *30*, 9–15. https://doi.org/10.1055/s-2008-1038768

Armstrong, R. B. (1984). Mechanicsms of Exercise-Induced Delayed Onset Muscle Soreness: a brief review.

Barden, J. M., & Kell, R. T. (2009). Relationships between stroke parameters and critical swimming speed in a sprint interval training set. *Journal of Sports Sciences*, *27*(912585618), 227–235. https://doi.org/10.1080/02640410802475205

Barnes, K. R., & Kilding, A. E. (2015). Running economy : measurement , norms , and determining factors, 1–15. https://doi.org/10.1186/s40798-015-0007-y

Beaufort, F. (1805). Beaufort Wind Scale. Retrieved from https://www.weather.gov/mfl/beaufort

Bland, J. M., & Altman, D. G. (2010). Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies*, *47*(8), 931–936. https://doi.org/10.1016/j.ijnurstu.2009.10.001

Broxterman, R. M., Ade, C. J., Craig, J. C., Wilcox, S. L., Schlup, S. J., & Barstow, T. J. (2014). The relationship between critical speed and the respiratory compensation point: Coincidence or equivalence. *European Journal of Sport Science*, *15*(October 2014), 1–9. https://doi.org/10.1080/17461391.2014.966764

Broxterman, R. M., Ade, C. J., Poole, D. C., Harms, C. A., & Barstow, T. J. (2013). A single test for the determination of parameters of the speed – time relationship for running. *Respiratory Physiology & Neurobiology*, *185*(2), 380–385. https://doi.org/10.1016/j.resp.2012.08.024

Burnley, M. (2009). Estimation of critical torque using intermittent isometric maximal voluntary contractions of the quadriceps in humans. *Journal of Applied Physiology*, *106*, 975–983.

Burnley, M., Doust, J. H., & Vanhatalo, A. (2006). A 3-min all-out test to determine peak oxygen uptake and the maximal steady state. *Medicine and Science in Sports and Exercise*, *38*(11), 1995–2003. https://doi.org/10.1249/01.mss.0000232024.06114.a6

Caspersen, C. J., & Christenson, G. M. (1985). Physical Activity , Exercise , and Physical Fitness : Definitions and Distinctions for Health-Related Research, (April).

Castellani, J. W., & Young, A. J. (2012). Health and performance challenges during sports training and competition in cold weather, 1–6. https://doi.org/10.1136/bjsports-2012-091260

Cheung, K., Hume, P. A., & Maxwell, L. (2003). Treatment Strategies and Performance Factors, *33*(2), 145–164.

Clark, I. E., West, B. M., Reynolds, S. K., Murray, S. R., & Pettitt, R. W. (2013). Applying the critical velocity model for an off-season interval training program. *Journal of Strength and Conditioning Researchournal of S*, *27*(12), 3335–3341.

Dekerle, J., & Paterson, J. (2016). Muscle Fatigue When Swimming Intermittently Above and Below Critical Speed. *International Journal of Sports Physiology and Performance*, *11*(Figure 1), 602–607.

Di Prampero, P. E. (1999). The concept of critical velocity: A brief analysis. *European Journal of Applied Physiology and Occupational Physiology*, *80*(2), 162–164. https://doi.org/10.1007/s004210050574

Faulkner, J., White, T., & Markley, J. (1981). The 1979 Canadian ski marathon: a natural experiment in hypothermia. *Exercise in Health and Disease*, 184–195.

Gaesser, G. A., & Poole, D. C. (1996). The slow component of oxygen uptake kinetics in humans. *Exercise and Sport Sciences Reviews*, *24*, 35–71.

Galbraith, A., Hopker, J., Lelliott, S., Diddams, L., & Passfield, L. (2014). A single-visit field test of critical speed. *International Journal of Sports Physiology and Performance*, *9*(6), 931–935. https://doi.org/10.1123/ijspp.2013-0507

Helou, N. El, Tafflet, M., Berthelot, G., Tolaini, J., Marc, A., Guillaume, M., & Hausswirth, C. (2012). Impact of Environmental Parameters on Marathon Running Performance, *7*(5), 1–9. https://doi.org/10.1371/journal.pone.0037407

Hill, D. W. (1993). The critical power concept: a review. *Sports Medicine*, *16*, 237–254.

IAAF. (2017). Rules 2016-2017.

Jones, A. M., & Carter, H. (2000). The effect of endurance training on parameters of aerobic fitness. *Sports Medicine (Auckland, N.Z.)*, *29*(6), 373–386. https://doi.org/10.2165/00007256-200029060-00001

June, B., Magal, M., Hester, M., Ivey, M. A., Eschbach, L. C., Bunn, J. A., … Florida, C. (2011). The Importance of Preseason Physical Fitness Testing in Professional Soccer Players in Qatar, 2011.

Kenttä, G., & Hassmén, P. (1998). Overtraining and Recovery A Conceptual Model. *Sports Medicine*, *26*(1), 1–16.

Kramer, M., Clark, I. E., Jamnick, N., Strom, C., & Pettitt, R. W. (2018). Normative Data For Critical Speed and D’ For High-Level Male Rugby Players, *32*(3), 783–789.

Landis, J. R., & Koch, G. G. (2008). The Measurement of Observer Agreement for Categorical Data. *International Biometric Society Stable*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Maryn, N., King, K., & Penry, J. (n.d.). A Comparison of Three Methods to Determine Critical Running Velocity in Healthy Adults.

Morris, J. G., Nevill, M. E., Lakomy, H. K. A., Nicholas, C., Williams, C., Nevill, M. E., … Williams, C. (2017). Effect of a hot environment on performance of prolonged , intermittent , high-intensity shuttle running E V ect of a hot environment on performance of prolonged , intermittent , high-intensity shuttle running, *414*(October). https://doi.org/10.1080/026404198366489

Nimmerichter, A., Steindl, M., & Williams, C. (2015). Reliability of the Single-Visit Field Test of Critical Speed in Trained and Untrained Adolescents. *Sports*, *3*(4), 358–368. https://doi.org/10.3390/sports3040358

Ortega, F. B., Ruiz, J. R., Castillo, M. J., & Sjöström, M. (2008). Physical fitness in childhood and adolescence: a powerful marker of health. *International Journal of Obesity*, *32*(1), 1–11. https://doi.org/10.1038/sj.ijo.0803774

Pettitt, R. W., Jamnick, N., & Clark, I. E. (2012). 3-min All-out Exercise Test for Running. *Int J Sports Med*, *33*, 426–431.

Pollock, M. L., Gaesser, G. A., Butcher, J. D., Despres, J. P., Dishman, R. K., Franklin, B. A., & Garber, C. E. (2015). ACSM Position Stand : The Recommended Quantity and Quality of Exercise for Developing and Maintaining Cardiorespiratory and Muscular Fitness , and Flexibility in Healthy Adults, *30*(June 1998), 1–28.

Poole, D. C., Burnley, M., Vanhatalo, A., Rossiter, H. B., & Jones, A. M. (2016). Critical Power : An Important Fatigue Threshold in Exercise Physiology. *Medicine and Science in Sports and Exercise*, 2320–2334. https://doi.org/10.1249/MSS.0000000000000939

Shete, A. N. (2014). A Study of VO 2 Max and Body Fat Percentage in Female Athletes. *Journal of Clinical and Diagnostic Research*, *8*(12), BC01-3. https://doi.org/10.7860/JCDR/2014/10896.5329

Skiba, P. F., Chidnok, W., Vanhatalo, A., & Jones, A. M. (2012). Modeling the expenditure and reconstitution of work capacity above critical power. *Medicine and Science in Sports and Exercise*, *44*(8), 1526–1532. https://doi.org/10.1249/MSS.0b013e3182517a80

Solomonson, A. A., Dicks, N. D., Kerr, W. J., & Petitt, R. W. (2016). Influenece of Load Carriage on High-Intensity Running Performance Estimation. *Journal of Strenght and Conditioning*, *30*(5), 19–21. https://doi.org/10.1519/JSC.0000000000001209

Teo, W., Newton, M. J., & McGuigan, M. R. (2011). Circadian rhythms in exercise performance : Implications for hormonal and muscular adaptation. *Journal of Sports Science and Medicine*, *10*(December), 600–606.

Thomas, D., Fernhall, B., & Granat, H. (1999). Changes in running economy during a 5-km run in trained men and women runners. *Journal of Strength and Conditioning Research*, *13*(2), 162–167.

Tyler, C. (2008). The Effect of Ambient Temperature on the Reliability of a Preloaded Treadmill Time-Trial, 812–816. https://doi.org/10.1055/s-2007-989321

Vanhatalo, A., Doust, J. H., & Burnley, M. (2007). Determination of critical power using a 3-min all-out cycling test. *Medicine and Science in Sports and Exercise*, *39*(3), 548–555. https://doi.org/10.1249/mss.0b013e31802dd3e6

Vanhatalo, A., Fulford, J., DiMenna, F. J., & Jones, A. M. (2010). Influence of hyperoxia on muscle metabolic responses and the power-duration relationship during severe-intensity exercise in humans: a ^{31} P magnetic resonance spectroscopy study. *Experimental Physiology*, *95*(4), 528–540. https://doi.org/10.1113/expphysiol.2009.050500

Warburton, D. E. R., Nicol, C. W., & Bredin, S. S. D. (2006). Health benefits of physical activity: the evidence. *CMAJ : Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, *174*(6), 801–9. https://doi.org/10.1503/cmaj.051351

Weir, S. F. J. P. (1997). Relationship of critical velocity to marathon running performance, 274–278.

APPENDICES

Appendix One

Participant Screening Questionnaire

____________________ _________________________

Date Subject ID

1. Please indicate whether you are male or female	Male	Female
2. What is your age?	___ years	
3. Do you regularly participate in cardiovascular physical activity a minimum of three times a week on average?	Yes	No
4. Please list the type(s) of cardiovascular physical activity that you typically participate in, the number of times you complete it per week, and the average duration. Type of Activity: Number of times per week: Average length of each session: __________ __________ __________ __________ __________ __________ __________ __________ __________ __________ __________ __________		
5. Do you consider yourself an aerobically fit individual as defined by being able to complete a run of at least three miles at a 10-minute per mile pace?	Yes	No
6. Do you currently have any injury or illness that restricts your physical activity?	Yes	No
7. Has any medical provider recommended restrictions on your physical activity that would include not running as fast as possible for three minutes?	Yes	No

8. Have you ever had <u>any</u> of the following? a. heart attack b. heart failure c. cardiac arrhythmia d. known heart murmur e. congenital heart disease f. any heart surgery g. coronary angioplasty h. heart palpitations	Yes	No
9. Have you <u>ever</u> experienced any of the following? a. chest pain with mild exertion b. dizziness, fainting, or blackouts with mild exertion c. unusual fatigue or shortness of breath during usual activities	Yes	No
10. Have you <u>ever</u> been prescribed heart medications?	Yes	No
11. Do you smoke?	Yes	No
12. Have you been diagnosed with a blood pressure greater than 140/90?	Yes	No
13. Do you take blood pressure medication?	Yes	No
14. Are you diabetic or do you take medicine to control your blood sugar?	Yes	No
15. Have you been diagnosed with high cholesterol (>200)?	Yes	No
16. Do you have a close blood relative who had a heart attack before age 55 (father/brother) or age 65 (mother/sister)?	Yes	No

Research Team: Review participant's screening questionnaire. Continue on to assessment of BMI only if <u>all of the following apply</u>:

☐ *Age is 18-35 years* ☐ *YES on questions 3 and 5* ☐ *NO on questions 6 through 16*

BMI Assessment:

Height:	_______ m
Weight:	______ kg

Research Team: Calculate BMI= (Weight)/(Height)^2 = ______________

If participant PASSES screening process (BMI < 30.0 kg/m^2), have participant complete the following:

---

1. Please provide your preferred method of communication?

   Phone: ____________________ OR Email:_________________________

2. Please provide your expected availability by providing the best days and time frames for you to complete the testing sessions. (Circle all that apply)

Sunday:
Anytime / Not Available / 8-11 AM / 11-2 PM / 2-5 PM / 5-8 PM / Other: ___________

Monday:
Anytime / Not Available / 8-11 AM / 11-2 PM / 2-5 PM / 5-8 PM / Other: ___________

Tuesday:
Anytime / Not Available / 8-11 AM / 11-2 PM / 2-5 PM / 5-8 PM / Other: ___________

Wednesday:
Anytime / Not Available / 8-11 AM / 11-2 PM / 2-5 PM / 5-8 PM / Other: ___________

Thursday:
Anytime / Not Available / 8-11 AM / 11-2 PM / 2-5 PM / 5-8 PM / Other: ___________

Friday:
Anytime / Not Available / 8-11 AM / 11-2 PM / 2-5 PM / 5-8 PM / Other: ___________

Saturday:
Anytime / Not Available / 8-11 AM / 11-2 PM / 2-5 PM / 5-8 PM / Other: ___________

Appendix Two

Data Collection

____________ Rater Code   ________________ Participant Code   _____________ Date   _____________ Session #

Record each time split at every cone

Split #	Time (sec)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	

Calculating CS:

Input data from last two complete splits

(Split 1 time __________ + Split 2 time ___________)/2

= 50m/______________=______________ CS in m/s

Appendix Three

Test-Retest Environmental Conditions

Subject	Test 1			Test 2			
	Wind Speed (mph)	Temp (°F)	Time	Wind Speed (mph)	Temp (°F)	Time	Time Between (hrs)
1	6.5	76	4:00 PM	2.2	75	2:00 PM	118
2	3.0	73	1:30 PM	4.4	68	1:30 PM	120
3	2.8	72	1:30 PM	3.1	68	1:30 PM	120
4	6.7	76	4:00 PM				
5	4.5	74	3:00 PM	3.6	71	3:00 PM	72
6	2.5	59	8:00 AM	1.4	65	9:00 AM	97
7	3.5	59	8:00 AM	2.2	64	9:00 AM	97
8	2.5	59	8:00 AM	1.0	69	8:00 AM	72
9	3.0	60	8:00 AM	1.0	69	8:00 AM	72
10	2.2	68	8:30 AM	1.0	69	8:30 AM	72
11	2.4	68	8:30 AM				
12	2.2	68	8:30 AM				
13	1.7	64	9:30 AM	0.7	72	9:30 AM	96
14	2.2	65	9:30 AM	2.0	72	9:30 AM	96
15	1.2	72	8:00 AM				
16	1.6	72	8:00 AM				