AN ABSTRACT OF THE DISSERTATION OF

Jason T. Penry for the degree of Doctor of Philosophy in Exercise and Sport Science presented on December 12, 2008.
Title: Validity and Reliability Analysis of Cooper’s 12-Minute Run and the Multistage Shuttle Run in Healthy Adults.

Abstract approved:

________________________________________
Anthony R. Wilcox

Field tests are often the most practical method to assess aerobic fitness, but demonstrate greater error variability than laboratory tests. Researchers can improve field tests by identifying factors that contribute to systematic error in VO2max estimation. PURPOSE: To examine the validity and reliability evidence of two field tests of aerobic fitness: Cooper’s 12-minute run (12MR) and the multistage shuttle run (MSR). METHODS: Sixty participants (mean age = 21.8 ± 3.6y), completed three trials (occasion) of each field test (instrument) for a total of six test trials. To estimate overall reliability and evaluate possible sources of error in the field tests, a psychometric statistical tool called generalizability study (G-study) was employed. This analysis utilized a two-random facet design (occasion and instrument) in a completely crossed ANOVA. In addition, criterion VO2max was assessed in a subgroup of volunteers (n = 21) via an incremental treadmill run and expired gas analysis (TR). Each participant completed the study within a six-week period. RESULTS: G-study analysis of the two field tests returned a high reliability coefficient (φ = 0.96), with the largest amount of systematic error variance (4.3%) attributable to an interaction between participants and test occasions. This mild interaction suggests certain test
participants demonstrated larger error variability across test occasions than other participants. The MSR predicted VO\textsubscript{2max} values lower than those measured in the laboratory setting (p < 0.01; paired t-tests), while 12MR and TR scores were not different (p > 0.05). The 12MR underestimated VO\textsubscript{2max} values at lower aerobic fitness levels and overestimated VO\textsubscript{2max} values in individuals demonstrating greater aerobic fitness, which was not observed in the MSR data. CONCLUSIONS: These results suggest high reliability for VO\textsubscript{2max} field tests in young, healthy individuals. However, test administrators must use caution when attempting to use field test data to predict criterion VO\textsubscript{2max} scores. While test participants can be expected to attain MSR scores significantly lower than the criterion value, the consistent mean bias across VO\textsubscript{2max} values makes the MSR a more useful test when comparing test participants.
Validity and Reliability Analysis of Cooper’s 12-Minute Run and the Multistage Shuttle Run in Healthy Adults

by

Jason T. Penry

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Presented December 12, 2008
Commencement June 2009

APPROVED:

_____________________________________________________________________

Major Professor, representing Exercise and Sport Science

_____________________________________________________________________

Chair of the Department of Nutrition and Exercise Science

_____________________________________________________________________

Dean of the Graduate School

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

_____________________________________________________________________

Jason T. Penry, Author
ACKNOWLEDGEMENTS

I would like to thank:

– My wife Julie, whose support though this process has been without measure.

– Dr. Tony Wilcox, for his guidance and mentorship.

– Dr. Joonkoo Yun, Dr. Melinda Manore, Dr. Jeff Widrick, and Dr. Dwaine Plaza, whose individual contributions helped to bring this document to fruition.

– Dr. Janet Peterson, Dr. Garry Killgore, Cassie Hansen, Nelson Helland, and Lance Reem, who assisted with data collection at Linfield College.

– Dr. Peterson and Stasi Kasianchuk, for helping recruit study participants.

– Every study participant who woke up early in the morning and worked hard.

– My friends and family, who will (still) invite me to Thanksgiving dinner next year.
CONTRIBUTION OF AUTHORS

Dr. Tony Wilcox and Dr. Joonkoo Yun were involved in project design and draft revision for this document. Dr. Yun assisted with the analysis and interpretation of the G-study data.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Methods</td>
<td>4</td>
</tr>
<tr>
<td>Experimental approach to problem</td>
<td>4</td>
</tr>
<tr>
<td>Participants</td>
<td>4</td>
</tr>
<tr>
<td>Measures</td>
<td>5</td>
</tr>
<tr>
<td>Procedures</td>
<td>7</td>
</tr>
<tr>
<td>Statistical analyses</td>
<td>9</td>
</tr>
<tr>
<td>Results</td>
<td>11</td>
</tr>
<tr>
<td>Discussion</td>
<td>13</td>
</tr>
<tr>
<td>Practical applications</td>
<td>18</td>
</tr>
<tr>
<td>References</td>
<td>19</td>
</tr>
<tr>
<td>Appendices</td>
<td>28</td>
</tr>
<tr>
<td>Appendix A (IRB protocol)</td>
<td>28</td>
</tr>
<tr>
<td>Appendix B (consent forms and data collection sheets)</td>
<td>32</td>
</tr>
<tr>
<td>Appendix C (raw data)</td>
<td>41</td>
</tr>
<tr>
<td>Appendix D (data collection journal)</td>
<td>51</td>
</tr>
<tr>
<td>Appendix E (proposal document)</td>
<td>59</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Descriptive data</td>
<td>23</td>
</tr>
<tr>
<td>2.</td>
<td>Correlation matrix for the 12MR, MSR, and TR</td>
<td>23</td>
</tr>
<tr>
<td>3.</td>
<td>Results of the 2-facet G-study</td>
<td>23</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bland-Altman plot for the 12MR vs. the TR</td>
<td>24</td>
</tr>
<tr>
<td>2. Bland-Altman plot for the MSR vs. the TR</td>
<td>25</td>
</tr>
<tr>
<td>3. Bland-Altman plots for trials of the 12MR</td>
<td>26</td>
</tr>
<tr>
<td>4. Bland-Altman plots for trials of the MSR</td>
<td>27</td>
</tr>
</tbody>
</table>
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Term</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body mass index</td>
<td>BMI</td>
</tr>
<tr>
<td>Cooper’s 12-minute run</td>
<td>12MR</td>
</tr>
<tr>
<td>Decision study (follow-up to G-study)</td>
<td>D-study</td>
</tr>
<tr>
<td>Generalizability theory</td>
<td>G-theory</td>
</tr>
<tr>
<td>Generalizability study</td>
<td>G-study</td>
</tr>
<tr>
<td>Incremental treadmill run (criterion measure)</td>
<td>TR</td>
</tr>
<tr>
<td>Instrument facet</td>
<td>i</td>
</tr>
<tr>
<td>Maximal aerobic power</td>
<td>VO$_{2\text{max}}$</td>
</tr>
<tr>
<td>Multistage shuttle run</td>
<td>MSR</td>
</tr>
<tr>
<td>Occasion facet</td>
<td>o</td>
</tr>
<tr>
<td>Oregon State University</td>
<td>OSU</td>
</tr>
<tr>
<td>Participant facet</td>
<td>p</td>
</tr>
<tr>
<td>Pearson’s product-moment coefficient</td>
<td>PMCC</td>
</tr>
<tr>
<td>Random error</td>
<td>e</td>
</tr>
<tr>
<td>Rating of perceived exertion</td>
<td>RPE</td>
</tr>
<tr>
<td>Respiratory exchange ratio</td>
<td>RER</td>
</tr>
</tbody>
</table>
Validity and Reliability Analysis of Cooper’s 12-Minute Run and the Multistage Shuttle Run in Healthy Adults

The physiological parameter most closely linked to human endurance performance is maximal aerobic power, or VO$_{2\text{max}}$ (35). As such, accurate tests of this attribute are important, as the data are critical to prescribing physical activity in individuals of all ages. VO$_{2\text{max}}$ is most accurately measured in a laboratory setting by collecting expired gases, and demonstrates excellent reliability ($r = 0.95$) under such conditions (32). However, this technique requires sophisticated equipment and testing expertise. Access to such resources is not always available or feasible, and a need exists for valid and reliable tests that can be used by a variety of test administrators in a non-laboratory environment.

To address this demand, researchers have developed a number of tests to quantify VO$_{2\text{max}}$ outside the laboratory (5, 15, 26). These “field” tests are generally more variable than their laboratory counterparts, as they (a) estimate VO$_{2\text{max}}$ from the performance of a physical task, rather than measuring oxygen uptake, and (b) are conducted in a less-controlled environment than the laboratory. As such, a test administrator may doubt that a field test score is indicative of the true VO$_{2\text{max}}$ score for an individual. This uncertainty is reinforced by the discrepancy in validity coefficients across testing populations. Two of the most commonly used VO$_{2\text{max}}$ field tests in adults (26), Cooper’s 12- minute run (12MR) and the multistage shuttle run (MSR), demonstrate a high Pearson product-moment correlation coefficient (PMCC) with a criterion measure in aerobically fit populations ($r > 0.90$) (1, 5), but are less closely related ($r < 0.75$) to the criterion across the complete range of aerobic fitness levels.
(23, 28, 31). To collect accurate data, it is vital that test administrators understand and address the factors that increase score variability in heterogeneous populations.

A well-known source of test score variability is the day or occasion on which a participant completes a test. Resistance to this systematic error is often reported as “good” test-retest reliability. PMCC has been widely used to investigate such reliability evidence for the 12MR and MSR, with PMCC reported as high as 0.94 (8) for the 12MR and 0.96 (14) for the MSR. However, recent research suggests that the PMCC overestimates the agreement between consecutive trials of the MSR, as Lamb and Rogers (14) demonstrated that a high PMCC (r > 0.94) could mask variability in test scores that invalidated the MSR as a predictor of VO_{2max}. Thus, the PMCC may fail to accurately describe field test reliability, and an alternative measure of score variability may prove useful when examining reliability in such tests.

Accurate assessment of VO_{2max} field test score variability requires an analysis sensitive to the multifaceted relationships that exist between human participants and such field tests. A statistical tool that fulfills this requirement is generalizability theory (G-theory) (6). While classical test theory partitions score variability into “treatment” and “error” components, G-theory uses multiple repetitions of a treatment under selected conditions (called facets) to further partition the error variance into systematic error (attributable to each selected condition) and random error. Commonly used facets include test trial and test administrator, but a G-theory analysis (a G-study) can be designed to examine any test condition a researcher might deem important for investigation.
G-theory is specifically helpful to researchers attempting to describe and improve measurement tools used in exercise science. As such, it has been used to examine variability within tests of body composition (13, 20, 33), tests of isometric force (27), joint range of motion (22), and motor performance assessment (16, 21, 30). The ability of G-theory to describe the major contributors to error variance within tests also allows researchers to make specific recommendations to reduce this error and thus improve test protocol. For example, when Roebroeck et al. (27) determined test occasion was a large contributor to isometric force measurement error, they suggested researchers collect data on multiple occasions to account for this systematic error. Although no attempt has yet been made to apply G-theory to field tests of VO$_{2\max}$, this lack of application does not preclude the potential utility of G-theory in refining field tests of VO$_{2\max}$, as the work by Lamb and Rogers (14) evidences.

The purpose of the present study was to investigate the relative variability of the 12MR and MSR across test occasions via G-study. Criterion-related validity of the two field tests was also addressed. Subsequent examination of this G-study and the associated criterion-related validity measures should help test administrators collect more meaningful 12MR and MSR data from individuals of all aerobic fitness levels.
METHODS

Experimental approach to problem

The overall reliability of the 12MR and MSR was described via a 2-facet G-study model. This construct focused on the relative contribution of test instrument and test occasion to the total variability in test scores. Test instrument (23) and test occasion (14) were selected as facets within the G-study model due to the potential for these two factors to influence a participant’s field test score.

To best assess the validity of the 12MR and the MSR, it was important to compare the means of the two tests to the TR, as well as the correlation between each field test and the TR. Under-prediction of criterion VO_{2max} values via MSR appears widespread (4, 9, 11, 18), making it particularly important to assess mean bias in the current sample. Bland-Altman plots of the field test data were also constructed to look for trends in score variability over the range of VO_{2max} values. Establishing the validity of the two field tests in such a manner was necessary for the application of these data outside of the sample population, as it has been previously suggested that the validity of a particular field test may vary depending on the aerobic fitness of an individual (28).

Participants

Participants (n = 60; 33 females, 28 males) were recruited from the campuses of two higher education institutions in Oregon. The participants were contacted through in-class announcements, e-mail, and word-of-mouth advertising. Individuals of all aerobic fitness levels were encouraged to participate in the study. All participants were aged between 18 and 33 years (mean age = 21.8 ± 3.6) and did not
possess more than a single risk factor for cardiovascular disease (ACSM; see Appendix B). Physical characteristics for all test participants are presented in Table 1.

**Measures**

*Cooper’s 12-minute run (12MR)* – Participants were instructed to complete as many laps as possible on an all-weather, outdoor track during the 12-minute test period. Emphasis was placed on pacing oneself throughout the duration of the test. The test administrator counted the laps completed during the 12-minute test period, while calling out the time elapsed at 3, 6, and 9 minutes and verbally encouraging the participants. At the end of the 12-minute period, the test administrator called to the participants to “stop”, and a measuring wheel was used to determine the fraction of the last lap completed by each participant. This distance was added to the distance determined by the number of laps completed to give the total distance covered during the test. Cooper’s (5) standardized equation was then used to convert the distance run to an estimate of VO\(_{2\text{max}}\).

*Multistage 20-meter shuttle run (MSR)* – In a protocol first standardized by Leger et al. (15), participants ran a back-and-forth course (“shuttles”) between two rows of cones placed 20 meters apart from one another. The speed of the run started at 8.5 km/hr, and increased by 0.5 km/hr every minute. Participant speed was governed by a recorded, audible “beep” that sounded each time the participant was expected to reach a cone to complete a shuttle. The number of shuttles per stage was coordinated with the speed of the beeps so that each stage was approximately a minute in length. Participants received verbal encouragement throughout the duration of the test. When a participant failed to complete two successive shuttles in the allotted time between
“beeps”, the test was terminated for that participant. Estimated VO$_{2\text{max}}$ was determined from a printed table by using the number of shuttles completed (25).

Incremental treadmill test (TR) – A sub-sample of study participants (n = 21) volunteered to complete a TR to directly assess the criterion-related validity of the 12MR and MST. This test was conducted using a ParvoMedics TrueMax 2400 metabolic cart to measure oxygen consumption. During these tests, participants first completed a 3-minute warm-up stage. The speed of the treadmill during this warm-up stage was constant (4.5 to 7.0 mph, depending on the relative fitness of the individual) and the treadmill incline was set at 0%. Following this warm-up period, the treadmill grade was increased to 3.0% for the first of successive 1-minute stages. At the completion of each stage, the treadmill speed was increased by 0.5 mph. When the participant reached a rating of perceived exertion (RPE) of 13-14 on the Borg scale, the test administrator stopped increasing treadmill speed and began increasing the grade of the treadmill by 1% with each new stage. When the participant indicated that he or she could not continue, the test was terminated. A test was determined to be a “true” maximal test if a participant reached two of the following three criteria: (a) a VO$_2$ plateau evidenced by an oxygen uptake difference of less than 2.1 mL/kg from the previous stage (10), (b) a respiratory exchange ratio (RER) of greater than 1.15 (10), or (c) a maximal heart rate within ±10 beats of the age-predicted maximum (24). One participant did not attain the requirements necessary to determine a maximal test on the first attempt, and voluntarily re-tested within a week of the first TR.
Procedures

Prior to inclusion in the study, participants completed an informed consent form approved by the OSU Institutional Review Board (Appendix A). The informed consent form provided participants with detailed information concerning the two types of field tests they would be asked to complete. It also instructed participants to refrain from “strenuous” exercise for the 24-hour period preceding each test, and to maintain the same activity level throughout the duration of the study. Volunteers (Table 1) who completed the optional incremental treadmill test (TR) read and signed a second informed consent form (Appendix B) prior to the treadmill evaluation.

Participants completed three trials of the two different field tests, for a total of six field test trials. Individuals rested at least a day between tests, which typically resulted in two to three test sessions each week. Each participant completed all six tests within a six-week (43-day) window, although most participants completed testing within a much shorter period of time (mean = 22.8 days ± 6.0). For those individuals completing the TR, the test was conducted after the completion of all six field tests and within the same six-week time window. This six-week period was chosen because it was thought to be short enough to minimize error from any changes in VO2max associated with a training effect. This assumption is supported by Carter et al. (3), who showed that an intense six-week endurance training intervention increased VO2max by only 3-4%. As test participants were asked to maintain their current activity level over the duration of the present study, little change was expected in the true VO2max values for these individuals.
Most participants (92%) completed the field test trials in groups of 2 to 6 (92%; Appendix D) and between the hours of 8 AM and 10 AM (92%; Appendix D). The individuals included in a particular group did not remain consistent, and varied depending on the daily schedules of the participants. The type of test administered to a group of participants on a particular occasion was determined randomly through the use of a computer program (12).

Participants performed the 12MR on an outdoor 400-meter synthetic track. 12MR testing was only conducted under “good” conditions; that is, participants tested under conditions no worse than very light precipitation and mild winds. Two different test administrators were used during the 12MR data collection. Although no research has shown 12MR scores to be responsive to different test administrators, the possibility for such an interaction exists. The present G-study would assign this variability to the participant-by-instrument interaction term (p*i) if a test administrator consistently affected 12MR scores across all occasions, or the three-way interaction term plus error (p*i*o, e) if different test administrators increased score variability across 12MR test occasions.

MSR data were collected on an indoor athletic floor surface. As opposed to the 12MR, the test administrator for the MSR was kept constant across all test occasions, as the MSR has been shown to be sensitive to differences in verbal feedback (19). Moreover, since the MSR requires the test administrator to terminate each participant trial individually, the use of multiple test administrators could be expected to inflate the non-systematic error variability (p*i*o, e) associated with the MSR test.
Statistical analyses

Descriptive statistics – Mean values and standard deviations were calculated for all demographic data (age, height, weight, body mass index, mean VO$_{2\text{max}}$ values) using the SPSS 15.0 for Windows statistical package (SPSS Inc., 2006).

Correlation analysis – PMCC were calculated between the TR and the highest test scores for the 12MR and MSR (n = 21). The value attained from the TR test is an “optimal” score – that is, it is assumed this is the highest score possible for the tested individual. As such, the highest field test scores were considered the best approximations of optimal field test scores, and thus most analogous to the VO$_{2\text{max}}$ values measured by the TR. A PMCC was also calculated between the highest scores on the 12MR and the MSR for the measured sample.

Student’s t-test analysis – The relative significance of the differences between the TR and the highest 12MR and MSR scores were assessed via paired, 2-tailed Student’s t-tests. Student’s t-tests were also used to compare field test scores of individuals completing the optional TR test to the field test scores of those individuals who did not complete the TR. In this latter comparison, an unequal variance methodology was used to address the problem of unequal sample sizes.

Bland-Altman analysis – Two Bland-Altman plots were constructed from the criterion data, using the best score for each participant from a particular field test and that individual’s TR score. The technique used to construct these plots was similar to that used by Bland and Altman (2). Microsoft Excel was used to plot the mean difference between field test scores and TR scores against the average of the field test and TR scores.
To examine differences between field test trials, similar Bland-Altman plots were constructed using the field test data sets. Score differences for the two test trials were plotted on the Y-axis, while the mean values for the two trials were plotted on the X-axis.

*G-theory analysis* – A two-facet (test instrument and test occasion) generalizability theory analysis was used to examine the sources of variability of the two field tests. As part of this analysis, a 2 x 3 repeated measure ANOVA was used to calculate the estimated variance components for each source of variability (participant, p; instrument, i; and occasion, o), as well as the interactions among variables (p*i, p*o, and i*o) and the three-way interaction plus random error (p*i*o, e). These estimated variance components were then used to calculate the percentage of total variance in the sample attributable to each facet and facet-interaction pair.

A follow-up decision-study (D-study) was conducted using the G-study estimated variance components and the method described by Shavelson and Webb (29). A single instrument, 2-occasion ϕ-coefficient was calculated to permit a better comparison of the present data with the results of the test-retest methodology (PMCC) used by previous researchers. All facets were treated as random and the technique for absolute decision D-studies was utilized.
RESULTS

The mean measured TR VO$_{2\text{max}}$ for participants was 49.9 mL/kg/min ± 6.2. This was greater than the highest mean 12MR (48.0 mL/kg/min ± 8.2; $p = 0.06$) and MSR score (46.9 mL/kg/min ± 6.6; $p < 0.01$) for the same individuals (Table 1). Despite this difference, PMCC showed a strong, positive relationship between participants’ scores on the TR and the 12MR ($r = 0.87$), and the TR and the MSR ($r = 0.86$). The 12MR and the MSR also appeared closely related, demonstrating a very strong, positive relationship ($r = 0.91$). A correlation matrix for the PMCC data can be found in Table 2.

Examination of the Bland-Altman plots for the criterion data revealed an inconsistent bias for the 12MR across aerobic fitness levels. Inspection of the mean differences between 12MR and TR scores showed the 12MR to underestimate VO$_{2\text{max}}$ for lower-scoring individuals, while overestimating VO$_{2\text{max}}$ in higher-scoring individuals (Figure 1). This inconsistency was not observed in the MSR data (Figure 2).

The estimated variance components from the G-study are presented in Table 3. Across all individuals, the largest contributor to score variability was participant (91.5%, $p$), followed by the participant-by-occasion interaction (4.3%, $p^*o$) and the three-way interaction and error component (3.8%, $p^*i^*o$, e). Negative variance components were assumed to be related to sampling error due to their relatively small magnitude (29), and were given a value of zero when calculating percentages of total variance (7). All other facets and facet interaction variables contributed less than 0.1% to the total variance in the data set. Finally, the φ-coefficient, which estimates of the
reliability of the G-study model (29), was very high ($\phi = 0.98$) for the present data.

The $\phi$-coefficient remained large when adjusted for a single test instrument and two test occasions ($\phi = 0.96$).
DISCUSSION

The results of this analysis indicate the 12MR and MSR are reliable tests in healthy adults between the ages of 18 and 35. This study goes beyond previous studies that have examined field tests of aerobic power, as it utilized a G-study design to compartmentalize a portion of the systematic error associated with these tests. As only 3.8% of score variability was not explained by the present model (p*i*o, e; see Table 3), it appears factors not included as facets in this analysis play a minor role in determining the reliability associated with field tests of aerobic power.

Despite returning a high reliability coefficient ($\phi = 0.96$), field tests do not appear to accurately estimate criterion $VO_{2\text{max}}$ values. This study is the first to use a Bland-Altman methodology to demonstrate a bias in 12MR $VO_{2\text{max}}$ scores toward the most aerobically fit individuals (Figure 1). Such a bias in favor of higher-scoring individuals was not observed in the MSR. As a result, the MSR may be more useful in heterogeneous test populations spanning a range of aerobic fitness levels, even though it routinely predicts $VO_{2\text{max}}$ scores lower than the criterion value (Table 1; Figure 2).

The high reliability coefficient generated as part of the present G-study agrees with test-retest PMCC values described previously for the 12MR (8, 17, 28) and the MST (9, 14, 25) in young, healthy individuals. This value is not congruent, however, with all studies that have examined this topic in a similar population (23, 28, 30). The present data may help to explain this phenomenon. In the current model, the only facet to contribute a notable amount of systematic error variance was the participant-by-occasion interaction (p*o, 4.3%; Table 3). This small, yet significant, p*o interaction is underscored by the negligible amount of relative variability contributed by the test
occasion facet (o) to the model (<0.1%). Taken together, these two findings suggest a few test participants demonstrated large score variability across test occasions, while other participants did not. Field test studies including a greater number of these “inconsistent” individuals in a sample population could be expected to demonstrate lower reliability values. Moreover, by identifying and managing these individuals separately, test administrators could increase the average predictive value of these field tests. Such treatment would help to address the prohibitively large MSR confidence intervals noted by Lamb and Rogers (14).

It is likely that the observed p*o score variance is related to non-physiological variables intrinsic to test participants (23). These variables may include psychological factors (28) or issues related to familiarity with the testing modality (23), and are likely to be amplified in individuals with lower aerobic fitness (28). Additionally, this phenomenon may be more pronounced in tests that require participants to pace themselves over the duration of a time goal than the length of a distance goal (18). There is moderate evidence for such a trend in the present data, as the 12MR demonstrated the smallest amount of test-retest variability in those individuals with the highest scores (Figure 3). However, this trend is much less obvious in the MSR data (Figure 4), and it is thus unlikely that factors mediated by aerobic fitness are the only contributors to the p*o interaction value. Differences in testing conditions may have also contributed to an increase in score variability, as Wanamaker (34) showed weaker 12MR test-retest reliability in smaller test groups. However, as group size fluctuated randomly in the present study, the resulting variability was instead expected to be part of the random error term (p*i*o, e; Table 3). As the relative magnitude of this term
was small (3.8%), the contribution of group size to variability in field test scores appears negligible in the present sample.

The small magnitude of unaccounted error variability also suggests participant gender has little effect on the reliability values associated with field test VO₂max scores. This lack of a gender interaction is mirrored in the criterion comparison data, as both males and females show similar patterns in field test scores relative to criterion values (12MR, Figure 1; MSR, Figure 2). These findings are in accordance with Safrit et al. (28), who suggested differing validity and reliability values between male and female populations are an artifact of absolute aerobic fitness levels (28). Test administrators should be confident that the 12MR and MSR will behave similarly in both male and female populations.

It is somewhat surprising that no facet that included test instrument (i, p*i, i*o) contributed more than 0.1% of the variance to the G-study model. This suggests that both the 12MR and MSR behaved similarly in respect to test variability, both alone (i) and across participants (p*i) or test occasions (i*o). Thus, it appears that neither the 12MR nor MSR plays a role in compartmentalizing variability in test scores. Put another way, the present G-study model showed the 12MR and MSR to generate a similar magnitude of test score variability across individuals. As a result, both the 12MR and MSR appear equally reliable when assessing maximal aerobic power in the field.

Despite the high criterion PMCC values (Table 2), the MSR significantly underestimated TR scores (p < 0.01). This finding confirmed previous data (4, 9, 11, 18) comparing MSR scores to a criterion. A similar negative bias was also observed (p
= 0.06) in the 12MR data, but it was not of sufficient magnitude to be labeled as significant by the present statistical analysis. This was not surprising, as the 12MR generally shows little mean bias when compared to a criterion measure (4, 11, 18).

A smaller mean bias would seem to make the 12MR a better tool for estimating \( VO_{2\text{max}} \) values in the field. However, the large bias at both high and low \( VO_{2\text{max}} \) values (Figure 1) renders the 12MR of lesser use in sample populations that span a broad range of aerobic fitness levels. Currently, it appears the 12MR can do little more than rank participants on an ordinal scale. Screening participants for a smaller range of \( VO_{2\text{max}} \) values may help to increase the utility of 12MR scores, but this strategy would not be feasible in a population where the test administrator has little foreknowledge of his/her test sample. If a test administrator has little experience with his/her test participants, it may be easiest to utilize the MSR to estimate \( VO_{2\text{max}} \), and expect the test to underestimate true score \( VO_{2\text{max}} \) by 3-7% across all individuals (4, 9, 11, 18; Figure 2).

Both the \( p^i \) and \( p^i*o, e \) terms were small for the current G-study model, suggesting the use of two test administrators during the 12MR had little effect on field test scores. This is of practical significance to exercise science professionals, as it shows test scores can be generalized across test administrators. Future research could confirm this finding by including test administrator as a facet within a subsequent G-study.

One weakness associated with the criterion data presented in this study reflects the method by which test participants were self-selected into the sub-sample that completed the TR. On average, participants included in the criterion assessment sub-
sample scored higher on the 12MR (p = 0.03) and MSR (p = 0.09) than those
individuals not included in the sub-sample. These differences were greatest between
the groups of female participants (Table 1). As a result, the conclusions derived from
the sub-sample may not be generalizable to the full sample population. It should be
noted, however, that the range of participant scores for the TR sample and the full
population were approximately equal. Future studies should attempt to include a larger
number of lesser aerobically fit individuals in the criterion assessment sub-sample,
although this may require a modification in the recruitment protocol.
PRACTICAL APPLICATIONS

Field tests show excellent reliability ($\varphi = 0.96$) in healthy individuals between the ages of 18 and 35. Test administrators should be confident that the 12MR and MSR will return similar VO$_{2\text{max}}$ values, and that repeated testing will not drastically shift participant scores. While small, the greatest amount of systematic error (4.3%) appears related to the interaction between test participant and test occasion. This effect may be due to dissimilarities in psychological factors across individuals, including an intrinsic motivation to return consistent maximal test performances. Familiarity with the testing modality may have also contributed to this minor interaction effect. Strategies to reduce this source of error could include (a) providing adequate motivation to all individuals so as to ensure optimal test performance, and (b) familiarizing less-active individuals with the proper techniques needed to achieve optimal test scores.

While field test scores appear stable across test instruments and test occasions, test administrators should use caution when predicting the criterion VO$_{2\text{max}}$ from 12MR and MSR scores. Test administrators should be particularly aware that the 12MR did not demonstrate a consistent bias across the full range of VO$_{2\text{max}}$ values, underestimating VO$_{2\text{max}}$ in participants with lower aerobic fitness and overestimating scores in participants with higher VO$_{2\text{max}}$ values. This bias toward more fit individuals was not observed in MSR data, although the MSR significantly underestimated TR VO$_{2\text{max}}$ ($p < 0.01$). Test administrators attempting to predict criterion scores from field test data may find it easier to use the MSR, as it consistently underestimated TR scores by $\sim 3$ mL/kg/min across all fitness levels.
REFERENCES


Table 1. Descriptive data. F = females, M = males. F_{all}, M_{all} = data for all participants; F_F, M_F = data for individuals who completed the field tests but no incremental treadmill test; F_{FTR}, M_{FTR} = data for individuals who completed the field tests and the incremental treadmill test. Field test scores marked † and ‡ were significantly different (p < 0.05). Bolded field test scores indicate a significant paired t-test value (p \leq 0.05) versus the TR score.

<table>
<thead>
<tr>
<th>n</th>
<th>age (years)</th>
<th>height (cm)</th>
<th>weight (kg)</th>
<th>BMI</th>
<th>12MR (ml/kg/min)</th>
<th>MSR (ml/kg/min)</th>
<th>TR (ml/kg/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F_{all} = 33</td>
<td>21.5 ± 3.7</td>
<td>164 ± 5.1</td>
<td>62.4 ± 7.8</td>
<td>23.2 ± 2.5</td>
<td>40.7 ± 7.4</td>
<td>40.1 ± 6.0</td>
<td>X</td>
</tr>
<tr>
<td>F_F = 21</td>
<td>20.0 ± 1.5</td>
<td>163 ± 5.7</td>
<td>62.7 ± 8.3</td>
<td>23.6 ± 2.5</td>
<td>38.1±5.3†</td>
<td>38.0±5.0‡</td>
<td>X</td>
</tr>
<tr>
<td>F_{FTR} = 12</td>
<td>23.9 ± 5.0</td>
<td>166 ± 3.4</td>
<td>61.8 ± 7.4</td>
<td>22.5 ± 2.5</td>
<td>45.2±8.6‡</td>
<td>43.8±5.7‡</td>
<td>46.7 ± 5.7</td>
</tr>
<tr>
<td>M_{all} = 27</td>
<td>22.1 ± 3.5</td>
<td>178 ± 6.1</td>
<td>80.3±12.8</td>
<td>25.4 ± 3.6</td>
<td>49.9 ± 6.6</td>
<td>50.4 ± 5.3</td>
<td>X</td>
</tr>
<tr>
<td>M_F = 18</td>
<td>21.7 ± 3.1</td>
<td>177 ± 4.6</td>
<td>81.1±12.4</td>
<td>25.7 ± 3.5</td>
<td>48.9 ± 6.7</td>
<td>50.1 ± 5.2</td>
<td>X</td>
</tr>
<tr>
<td>M_{FTR} = 9</td>
<td>22.9 ± 4.3</td>
<td>178 ± 8.7</td>
<td>78.6±14.2</td>
<td>24.7 ± 3.9</td>
<td>51.8 ± 6.3</td>
<td>51.1±5.6</td>
<td>54.0 ± 3.9</td>
</tr>
</tbody>
</table>

Table 2. Correlation matrix (n = 21) for VO_{2max} values from the 12-minute run (12MR), multistage shuttle run (MSR), and the criterion incremental treadmill test (TR).

<table>
<thead>
<tr>
<th></th>
<th>12MR</th>
<th>MSR</th>
<th>TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>12MR</td>
<td>X</td>
<td>0.91</td>
<td>0.87</td>
</tr>
<tr>
<td>MSR</td>
<td>0.91</td>
<td>X</td>
<td>0.86</td>
</tr>
<tr>
<td>TR</td>
<td>0.87</td>
<td>0.86</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 3. Results of the 2-facet G-study (n = 60) conducted as part of the present study. Values marked * were set to zero when calculating percent variance due to the negative value of the estimated variance component for that facet or facet interaction pair (Cronbach et al., 1972). EVC = estimated variance component.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>EVC</th>
<th>% total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants (p)</td>
<td>68.3652</td>
<td>91.5</td>
</tr>
<tr>
<td>Test Instrument (i)</td>
<td>5.0241 x 10^{-15}</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Occasions (o)</td>
<td>0.05956</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>p*i</td>
<td>-2.53 x 10^{-13}</td>
<td>0*</td>
</tr>
<tr>
<td>p*o</td>
<td>3.31808</td>
<td>4.3</td>
</tr>
<tr>
<td>i*o</td>
<td>-0.04395</td>
<td>0*</td>
</tr>
<tr>
<td>p<em>o</em>i, error</td>
<td>2.93569</td>
<td>3.8</td>
</tr>
</tbody>
</table>
Figure 1. Bland-Altman plot for Cooper’s 12-minute run (12MR) vs. the criterion measure. The dashed line represents the average difference from the mean score, while the dotted lines mark the edge of the 95% confidence interval. Open diamonds represent female participants, while closed squared represent male participants. Note that in both genders, 12MR systematically underestimates the criterion scores at lower values, while overestimating the criterion score at higher values.
Figure 2. Bland-Altman plot for the multistage shuttle run (MSR) vs. the criterion measure. The dashed line represents the average difference from the mean, while the dotted lines mark the edge of the 95% confidence interval. Open diamonds represent female participants, while closed squares represent male participants. Note that the MSR consistently under-predicts the criterion variable across the range of criterion values.
Figure 3. Bland-Altman plots comparing trials of the 12MR. The dashed line represents the average difference from the mean, while the dotted lines mark the edge of the 95% confidence interval. Open diamonds represent female participants, while closed squared represent male participants. Note the increased distribution of 12MR score differences at lower VO$_{2\text{max}}$ values.
Figure 4. Bland-Altman plots comparing trials of the MSR. The dashed line represents the average difference from the mean, while the dotted lines mark the edge of the 95% confidence interval. Open diamonds represent female participants, while closed squared represent male participants. Note the relatively consistent distribution of MSR score differences across the range of VO₂max values.
APPENDIX A

Institutional Review Board (IRB) protocol
1. **Brief description:**
   This study will attempt to discover the largest sources of participant score variability in field tests of aerobic power via application of a G-study experimental design and analysis. It is believed that field test aerobic power scores will vary depending on (1) the type of field test used to measure aerobic power, and (2) the occasion on which an individual is tested.

2. **Background and significance:**
   Many different types of field tests and testing conditions are employed to determine aerobic power in experimental participants. This study will statistically examine and determine which factors contribute the largest amount of participant score variability within two popular field tests of aerobic power: Cooper’s 12-minute run and the 20-meter multistage shuttle run. Furthermore, a subset of participants will complete a maximal incremental treadmill test to establish the validity of the two field tests. The conclusions of this study will allow researchers to better allocate resources so that more precise field measures of aerobic power can be made.

3. **Methods and procedures:**
   Participants will arrive at the testing location at a pre-scheduled time. Testing will take place in the Greater Portland or Corvallis areas. Prior to the initial test, individuals will complete a short health history questionnaire (appended to this application) designed to assess general cardiovascular and musculoskeletal health. In the event that an individual demonstrates more than one risk factor for coronary artery disease (as defined by the American College of Sports Medicine), that individual will not be permitted to take part in the study. Participants will also be asked to maintain their current activity level throughout the duration of the study.

   Participants will complete three trials of two different field tests of aerobic power. Protocols specific to each field test are described in detail below. Participants will be tested in groups of three to five, with the order of the three tests determined randomly. At least one day of rest will be scheduled following each test session, and all test sessions will be completed within a 4-week period. If an individual indicates he/she has greater than mild muscle or joint pain, that individual will not be allowed to participate in the testing for that day.

   A subgroup of 20 participant volunteers will also complete a maximal, incremental treadmill test (protocol also described below). Prior to beginning this seventh test, participants will complete a second informed consent form (appended to this application).

   Descriptions of each test follow below:

   **Cooper’s 12-minute run.** This test will be conducted on an indoor or outdoor track. In the event the test must be conducted on an outdoor track, every effort will be made to ensure that testing conditions are “good.” In the event that testing conditions are
“poor” (i.e. high winds, heavy precipitation), the test will be rescheduled for a later date.

During Cooper’s 12-minute run, individuals will be asked to complete as many laps as possible in the period of 12 minutes. After 12 minutes have elapsed, the participants will be asked to stop, and the distance covered on the last lap will be measured. Participants will be verbally encouraged throughout the duration of this test.

**Multistage 20-meter shuttle run.** This test will be conducted on a wood or synthetic gymnasium surface. Participants will be asked to run back-and-forth between two cones set 20 meters apart from one another. Running speed will be controlled by an audible “beep” from a calibrated audio track – each time a participant hears a “beep,” that individual should be reaching one of the cones. The speed of the “beeps” will increase each minute until the participant can no longer keep pace. When an individual does not reach the cone in time for the “beep” on two consecutive cones, the test will be terminated for that participant. Participants will be verbally encouraged throughout the duration of this test, which is expected to last between 8 and 15 minutes.

**Incremental treadmill test.** This test will be administered in the Oregon State University Human Performance Lab. This test will require participants to run on a treadmill for a total of 10-15 minutes while wearing a mask that will collect exhaled air. During the test, participants will run for 3 minutes at a prescribed “easy” pace (6.0 to 7.0 mph, depending on relative fitness). After the completion of these 3 minutes, the treadmill incline will be increase to 3%, and the running pace will be increased by 0.5 mph every minute until the participant feels he/she is running at a “somewhat hard” pace (13-14 on the Borg RPE scale). Once this pace has been reached, the incline of the treadmill will be increased by 1% each minute until the point at which the individual indicates that he/she cannot continue. At this point, the test will be terminated. Participants will be verbally encouraged throughout the duration of this test.

An incremental treadmill test will be considered to be “maximal” if the participant attains two of the following three criteria: (1) a VO2 plateau evidenced by an oxygen uptake difference of less than 2.1 mL/kg from the previous stage, (2) an RER higher than 1.15, and (3) a maximal heart rate within 10 beats per minute of the age-predicted maximum. If a participant does not achieve a maximal test, he/she will be asked to return to the laboratory for a second incremental treadmill test, provided that second test falls within the four-week testing window.

4. **Risks/benefit assessment:**

Participants will experience short-term fatigue when completing the aerobic power field tests or the maximal treadmill test. The fatigue is similar to that felt after running a half-mile race for the treadmill test and the multistage shuttle run, and similar to that following a 1.5 mile race for Cooper’s 12-minute run. There is a very remote chance that individuals may suffer a heart attack during these maximal efforts, although this will be a very low risk for the study participants,
since the pre-screening will have determined them to be physically active and apparently healthy.

Participants will receive several different estimates of their maximal aerobic power as a result of participating in this study. Those participants that choose to complete the incremental treadmill test will also receive a laboratory measure of their maximal aerobic power.

5. **Participant population:**
Approximately 50 to 75 participants are expected to be recruited over the life of this study. Participants are expected to be of 18-35 years of age, with no obvious physical ailments that would prevent them from completing the aerobic fitness tests.

6. **Subject identification and recruitment:**
Participants will be recruited via on-campus flyers, in-class announcements, and verbal recruitment. All individuals on the Oregon State campus and within the selected EXSS courses will have equal opportunity to participate in this study.

Participants will also be recruited from the Portland area via verbal announcements and flyer distribution in local physical activity groups.

Participants will be able to volunteer for the maximal incremental treadmill test until the data from 20 participants has been collected. Once this number of participants has been met, no more individuals will be tested using the treadmill test.

7. **Compensation:**
No compensation will be given to those individuals who participate in this study.

8. **Informed consent process:**
Informed consent will be obtained through each participant’s reading and subsequent signature of a detailed informed consent form. This consent will be obtained before beginning the set of field tests, as well as before the maximal incremental treadmill test. An example of both consent forms have been appended to this application.

9. **Anonymity or confidentiality:**
Participants’ data will only be identified with an assigned alphanumeric code. This code will be assigned sequentially, with the first participant receiving code P001, the second receiving P002, and so on. All data will be secured in a file cabinet within a locked office.
APPENDIX B

Consent forms and data collection sheets
INFORMED CONSENT DOCUMENT

Project Title: An assessment of variability in field tests of aerobic power.
Principal Investigator: Anthony Wilcox
Co-Investigator(s): Jason Penry, Joonkoo Yun

WHAT IS THE PURPOSE OF THIS STUDY?

You are being invited to take part in a research study designed to investigate the differences in several field tests of aerobic power, and to determine which factors may play a role in reducing the accuracy of these tests. It is believed that scores on field tests of aerobic power will vary depending on (1) the specific protocol used to determine aerobic power, and (2) the occasion on which the field test is administered. The information acquired from this study will clarify future research involving field tests of aerobic power, as it will allow researchers and other health professionals to use field tests to more accurately gather data about an individual’s aerobic fitness.

The investigators intend to publish these findings in a peer-reviewed journal and present these results at a professional conference in the near future. This study will also serve as the doctoral dissertation research for Jason Penry, one of the co-investigators named above.

WHAT IS THE PURPOSE OF THIS FORM?

This consent form gives you the information you will need to help you decide whether to be in the study or not. Please read the form carefully. You may ask any questions about the research, the possible risks and benefits, your rights as a volunteer, and anything else that is not clear. When all of your questions have been answered, you can decide if you want to be in this study or not.

WHY AM I BEING INVITED TO TAKE PART IN THIS STUDY?

You are being invited to take part in this study because you are an apparently healthy adult, aged 18-35.

WHAT WILL HAPPEN DURING THIS STUDY AND HOW LONG WILL IT TAKE?

During this study, you will participate in three repetitions of two field tests of aerobic power, for a total of six test sessions. Each test day will be followed by at least one day of rest (no test), and you will be asked to complete all tests within a four-week period.

You are asked to maintain your current activity level and refrain from strenuous activity for the period of 24 hours before each test.

Descriptions of each test follow below:

Cooper’s 12-minute run. This test will be conducted on an indoor or outdoor track in the Greater Portland or Corvallis area. In the event the test is conducted on an outdoor track, every effort will be made to ensure that testing conditions are “good.” In the event that testing conditions are “poor” (i.e. high winds, heavy precipitation), the test will be rescheduled for a later date.
During Cooper’s 12-minute run, you will be asked to run as many laps as possible in the period of 12 minutes. It is important that you pace yourself, and not run too quickly at the beginning of the test. The overall intensity for this test could be described as “somewhat hard” to “hard” throughout the test. After 12 minutes have elapsed, you will be asked to stop, and the distance covered on your last lap will be measured.

**Multistage 20-meter shuttle run.** This test will be conducted on a wood or synthetic gymnasium surface. You will be asked to run back-and-forth between two cones set 20 meters apart from one another. The speed at which you run will be controlled by an audible “beep” from a calibrated audio track – each time you hear a “beep,” you should be reaching one of the cones. The speed of the “beeps” will increase each minute until you can no longer keep pace. When you do not reach the cone in time for the “beep” on two consecutive cones, the test will be terminated. You can expect to be running for 8-15 minutes during this test, with the last 2-3 minutes at an intensity that could be described as “very hard.”

**Incremental treadmill test.** Participation in an incremental treadmill test to accompany this study is voluntary and will require you to complete an additional informed consent form. This test will take place in the Oregon State University Human Performance Laboratory, and will require you to run for 10-12 minutes on a treadmill while wearing a mask to collect the air you breathe out. The fatigue experienced following this test will be similar to that felt after completing a half-mile race.

**WHAT ARE THE RISKS OF THIS STUDY?**

You can expect to experience short-term fatigue when completing the aerobic power field tests. There is also a very remote chance that you may suffer a heart attack during a maximal effort on a treadmill. This is considered a low risk for you, since you are physically active and apparently healthy. In addition, every effort will be made to ensure that the areas in which the tests are conducted are free of obstacles that may cause injury. In the event of research-related injury, compensation for medical treatment is not provided by Oregon State University or the researchers.

**WHAT ARE THE BENEFITS OF THIS STUDY?**

You will receive information concerning your maximal aerobic power as a result of participating in this study. Moreover, in the future, other people might benefit from this study as it will allow physical education teachers, coaches, or researchers to better use field tests of aerobic power to collect data from their test participants.

**WILL I BE PAID FOR PARTICIPATING?**

You will not be paid for being in this research study.

**WHO WILL SEE THE INFORMATION I GIVE?**

The information you provide during this research study will be kept confidential to the extent permitted by law. To help protect your confidentiality, the data collected during this study will only be linked to your person by a numeric identification code.

If the results of this project are published, your identity will not be made public.
DO I HAVE A CHOICE TO BE IN THE STUDY?

If you decide to take part in the study, it should be because you really want to volunteer. You will not lose any benefits or rights you would normally have if you choose not to volunteer. You can stop at any time during the study and still keep the benefits and rights you had before volunteering. You will not be treated differently if you decide to stop taking part in the study. If you choose to withdraw from this project before it ends, the researchers may keep information collected about you and this information may be included in study reports.

WHAT IF I HAVE QUESTIONS?

If you have questions about this research project, please e-mail Tony Wilcox (anthony.wilcox@oregonstate.edu) or Jay Penry (penryj@onid.orst.edu).

If you have questions about your rights as a participant, please contact the Oregon State University Institutional Review Board (IRB) Human Protections Administrator, at (541) 737-4933 or by email at IRB@oregonstate.edu.

Your signature indicates that this research study has been explained to you, that your questions have been answered, and that you agree to take part in this study. You will receive a copy of this form.

Participant's Name (printed): __________________________________________________________

(Signature of Participant)  (Date)
INFORMED CONSENT DOCUMENT

Project Title: An assessment of variability in field tests of aerobic power.
Principal Investigator: Anthony Wilcox
Co-Investigator(s): Jason Penry, Joonkoo Yun

WHAT IS THE PURPOSE OF THIS STUDY?

You are being invited to take part in a research study designed to investigate the differences in several field tests of aerobic power, and to determine which factors may play a role in reducing the accuracy of these tests. The data from this maximal treadmill test will allow the study investigators to better understand the accuracy of the data collected from the aerobic power field tests that you previously completed. Moreover, the information acquired from this study will clarify future research involving field tests of aerobic power, as it will allow researchers and other health professionals to use field tests to more accurately gather data about an individual’s aerobic fitness.

The investigators intend to publish these findings in a peer-reviewed journal and present these results at a professional conference in the near future. This study will also serve as the doctoral dissertation research for Jason Penry, one of the co-investigators named above.

WHAT IS THE PURPOSE OF THIS FORM?

This consent form gives you the information you will need to help you decide whether to be in the study or not. Please read the form carefully. You may ask any questions about the research, the possible risks and benefits, your rights as a volunteer, and anything else that is not clear. When all of your questions have been answered, you can decide if you want to be in this study or not.

WHY AM I BEING INVITED TO TAKE PART IN THIS STUDY?

You are being invited to take part in this study because you are an apparently healthy adult, aged 18-35.

WHAT WILL HAPPEN DURING THIS STUDY AND HOW LONG WILL IT TAKE?

During this study, you will participate in a maximal incremental treadmill test.

You are asked to maintain your current activity level and refrain from strenuous activity for the period of 24 hours before the test.

You will participate in a maximal oxygen carrying capacity test (VO2max) on the treadmill. You will start the test with a 3-minute warm up of running at 6, 6.5 or 7 MPH (depending on your running experience) at 3% elevation. After the warm-up, the speed will increase by 0.5 mph each minute, until you reach the speed that you rate as being “somewhat hard”, which will be somewhere between 7.5-9 mph (again depending on your running experience). From this point, the treadmill will be elevated by 1% per minute with no further increases in speed, until you indicate that you are too fatigued to continue. The test usually takes 10-15 minutes with only the last few (3-5) minutes at a “hard” pace. During the test, you will be wearing a heart-rate monitor and a breathing valve through
which you can inhale room air and which directs your exhaled air to the metabolic cart, where your oxygen consumption will be determined.

The session will occur in the Human Performance Lab (Women’s Building, room 19, at Oregon State University), and will take approximately 30 minutes.

**WHAT ARE THE RISKS OF THIS STUDY?**

You can expect to experience short-term fatigue when completing the maximal treadmill test. This fatigue is similar to that felt after running a half-mile race. There is a very remote chance that you may suffer a heart attack during the maximal effort on the treadmill. This is considered a low risk for you, since you are physically active and are apparently healthy. In the event of research-related injury, compensation for medical treatment is not provided by Oregon State University or the researchers.

**WHAT ARE THE BENEFITS OF THIS STUDY?**

You will receive information concerning your maximal aerobic power as a result of participating in this study. Moreover, in the future, other people might benefit from this study as it will allow physical education teachers, coaches, or researchers to better use field tests of aerobic power to collect data from their test participants.

**WILL I BE PAID FOR PARTICIPATING?**

You will not be paid for being in this research study.

**WHO WILL SEE THE INFORMATION I GIVE?**

The information you provide during this research study will be kept confidential to the extent permitted by law. To help protect your confidentiality, the data collected during this study will only be linked to your person by a numeric identification code.

If the results of this project are published, your identity will not be made public.

**DO I HAVE A CHOICE TO BE IN THE STUDY?**

If you decide to take part in the study, it should be because you really want to volunteer. You will not lose any benefits or rights you would normally have if you choose not to volunteer. You can stop at any time during the study and still keep the benefits and rights you had before volunteering.

You will not be treated differently if you decide to stop taking part in the study. If you choose to withdraw from this project before it ends, the researchers may keep information collected about you and this information may be included in study reports.
WHAT IF I HAVE QUESTIONS?

If you have questions about this research project, please e-mail Tony Wilcox (anthony.wilcox@oregonstate.edu) or Jay Penry (penryj@onid.orst.edu).

If you have questions about your rights as a participant, please contact the Oregon State University Institutional Review Board (IRB) Human Protections Administrator, at (541) 737-4933 or by email at IRB@oregonstate.edu.

Your signature indicates that this research study has been explained to you, that your questions have been answered, and that you agree to take part in this study. You will receive a copy of this form.

Participant's Name (printed): __________________________________________________________

__________________________________________ _______________________________
(Signature of Participant)                (Date)
Date ______________  

**Health History Questionnaire**

Full Name: __________________________________  Date: ___/___/___  
Age: ______   Date of Birth: ___/___/___  I.D. #: _________  

The purpose of this questionnaire is to obtain information regarding your health prior to conducting physiological testing. Please answer all questions to the best of your knowledge. Circle the correct answers.

1. Do you have high blood pressure?    YES  NO  
2. Do you have high blood cholesterol?    YES  NO  
3. Do you currently have any muscle or joint pain?                   YES                NO  
   If yes, is this pain mild, moderate, or severe?       MILD     MODERATE     SEVERE  
3. Do you currently smoke?     YES  NO  
4. Are you a former smoker?     YES  NO  
   If so, when did you quit? ________________  
5. Have you ever had a heart attack?    YES  NO  
6. Have you ever had chest pain (angina)?    YES  NO  
7. Have any of your blood relatives had heart disease, heart surgery, or angina?  
   YES  NO  
   If so, what is the relation? __________What did they have?___________  
8. Are you diabetic?      YES  NO  
   If so, list medications taken.______________________________________  
9. Do you have any respiratory problems (Example: asthma, emphysema)?    YES  NO  
   If so, list them. ________________________________________________  
10. Have you had any recent illness, hospitalization, or surgical procedures?  
   YES  NO  
   If so, list them and when?________________________________________  
11. Are you currently taking any medications?   YES  NO  
   If so, list them. ________________________________________________  
12. Are you currently physically active?    YES  NO  
   Is so, how many minutes/week do you spend in moderate physical activity? ___________  
   What types of physical activity do you participate in?  

Please provide us with emergency contact information.

Name: __________________________  Home Phone: ________________  
Relation: _______________________  Work Phone: ________________
**Yo-Yo (Beep) Test Record Form**

Player Name: ________________________________

Age: _______ Sex: M F

Date of Test: ________________________________

<table>
<thead>
<tr>
<th>Level</th>
<th>Shuttle number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>3</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>4</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>5</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>6</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>7</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>8</td>
<td>1 2 3 4 5 6 7 8 9 10 11</td>
</tr>
<tr>
<td>9</td>
<td>1 2 3 4 5 6 7 8 9 10 11</td>
</tr>
<tr>
<td>10</td>
<td>1 2 3 4 5 6 7 8 9 10 11</td>
</tr>
<tr>
<td>11</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12</td>
</tr>
<tr>
<td>12</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12</td>
</tr>
<tr>
<td>13</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13</td>
</tr>
<tr>
<td>14</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13</td>
</tr>
<tr>
<td>15</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13</td>
</tr>
<tr>
<td>16</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14</td>
</tr>
<tr>
<td>17</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14</td>
</tr>
<tr>
<td>18</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
<tr>
<td>19</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
<tr>
<td>20</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
</tr>
<tr>
<td>21</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
</tr>
</tbody>
</table>
APPENDIX C

Raw data
<table>
<thead>
<tr>
<th>sub</th>
<th>sex</th>
<th>age</th>
<th>height</th>
<th>weight</th>
<th>days</th>
<th>pamin</th>
<th>12MR1</th>
<th>12MR2</th>
<th>12MR3</th>
<th>MSR1</th>
<th>MSR2</th>
<th>MSR3</th>
<th>TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>P001</td>
<td>1</td>
<td>22</td>
<td>180.7</td>
<td>79.1</td>
<td>29</td>
<td>90</td>
<td>2488.9</td>
<td>44.3</td>
<td>2468.1</td>
<td>43.88</td>
<td>2594.9</td>
<td>46.71</td>
<td>9</td>
</tr>
<tr>
<td>P002</td>
<td>2</td>
<td>18</td>
<td>156.2</td>
<td>54.5</td>
<td>27</td>
<td>180</td>
<td>2400.0</td>
<td>42.36</td>
<td>2509.3</td>
<td>44.8</td>
<td>2340.2</td>
<td>41.0</td>
<td>7</td>
</tr>
<tr>
<td>P003</td>
<td>2</td>
<td>19</td>
<td>167.6</td>
<td>75.0</td>
<td>22</td>
<td>450</td>
<td>2484.8</td>
<td>44.3</td>
<td>2452.3</td>
<td>43.52</td>
<td>2485.6</td>
<td>44.27</td>
<td>7</td>
</tr>
<tr>
<td>P004</td>
<td>2</td>
<td>19</td>
<td>170.8</td>
<td>73.2</td>
<td>22</td>
<td>540</td>
<td>2357.8</td>
<td>41.4</td>
<td>2000.0</td>
<td>33.42</td>
<td>2030.2</td>
<td>34.09</td>
<td>9</td>
</tr>
<tr>
<td>P005</td>
<td>2</td>
<td>21</td>
<td>160.0</td>
<td>66.8</td>
<td>22</td>
<td>360</td>
<td>2245.0</td>
<td>38.8</td>
<td>2257.0</td>
<td>39.16</td>
<td>2097.4</td>
<td>35.59</td>
<td>8</td>
</tr>
<tr>
<td>P006</td>
<td>2</td>
<td>19</td>
<td>165.7</td>
<td>81.8</td>
<td>22</td>
<td>90</td>
<td>1774.7</td>
<td>28.38</td>
<td>1746.0</td>
<td>27.74</td>
<td>1664.4</td>
<td>25.91</td>
<td>4</td>
</tr>
<tr>
<td>P007</td>
<td>2</td>
<td>19</td>
<td>177.8</td>
<td>98.2</td>
<td>27</td>
<td>600</td>
<td>2053.0</td>
<td>34.6</td>
<td>2432.4</td>
<td>43.08</td>
<td>2391.5</td>
<td>42.17</td>
<td>10</td>
</tr>
<tr>
<td>P008</td>
<td>2</td>
<td>18</td>
<td>165.7</td>
<td>81.8</td>
<td>22</td>
<td>90</td>
<td>2087.4</td>
<td>33.7</td>
<td>2071.6</td>
<td>34.79</td>
<td>2071.1</td>
<td>35.0</td>
<td>5</td>
</tr>
<tr>
<td>P009</td>
<td>2</td>
<td>19</td>
<td>160.0</td>
<td>55.0</td>
<td>22</td>
<td>270</td>
<td>2357.4</td>
<td>41.4</td>
<td>2269.5</td>
<td>39.44</td>
<td>2436.4</td>
<td>43.17</td>
<td>7</td>
</tr>
<tr>
<td>P010</td>
<td>1</td>
<td>18</td>
<td>177.8</td>
<td>98.2</td>
<td>27</td>
<td>600</td>
<td>2053.0</td>
<td>34.6</td>
<td>2432.4</td>
<td>43.08</td>
<td>2391.5</td>
<td>42.17</td>
<td>10</td>
</tr>
<tr>
<td>P011</td>
<td>2</td>
<td>21</td>
<td>168.9</td>
<td>66.4</td>
<td>22</td>
<td>90</td>
<td>2411.2</td>
<td>42.61</td>
<td>2395.0</td>
<td>42.24</td>
<td>2097.4</td>
<td>35.59</td>
<td>8</td>
</tr>
<tr>
<td>P012</td>
<td>2</td>
<td>19</td>
<td>170.2</td>
<td>68.2</td>
<td>22</td>
<td>90</td>
<td>1989.5</td>
<td>33.36</td>
<td>1613.5</td>
<td>24.78</td>
<td>1700.9</td>
<td>26.73</td>
<td>7</td>
</tr>
<tr>
<td>P013</td>
<td>1</td>
<td>19</td>
<td>175.3</td>
<td>70.0</td>
<td>27</td>
<td>60</td>
<td>2553.9</td>
<td>45.8</td>
<td>2548.0</td>
<td>45.66</td>
<td>2561.7</td>
<td>46.0</td>
<td>9</td>
</tr>
<tr>
<td>P014</td>
<td>2</td>
<td>18</td>
<td>161.9</td>
<td>66.8</td>
<td>22</td>
<td>720</td>
<td>2496.5</td>
<td>44.51</td>
<td>2572.4</td>
<td>46.21</td>
<td>2618.4</td>
<td>47.24</td>
<td>9</td>
</tr>
<tr>
<td>P015</td>
<td>2</td>
<td>20</td>
<td>157.5</td>
<td>55.9</td>
<td>22</td>
<td>150</td>
<td>1814.2</td>
<td>29.26</td>
<td>1784.8</td>
<td>28.61</td>
<td>1924.0</td>
<td>31.72</td>
<td>4</td>
</tr>
<tr>
<td>P016</td>
<td>2</td>
<td>20</td>
<td>160.0</td>
<td>59.1</td>
<td>22</td>
<td>90</td>
<td>2053.0</td>
<td>34.6</td>
<td>2432.4</td>
<td>43.08</td>
<td>2391.5</td>
<td>42.17</td>
<td>10</td>
</tr>
<tr>
<td>P017</td>
<td>2</td>
<td>20</td>
<td>167.6</td>
<td>67.3</td>
<td>22</td>
<td>240</td>
<td>1987.6</td>
<td>33.14</td>
<td>1613.5</td>
<td>24.78</td>
<td>1700.9</td>
<td>26.73</td>
<td>8</td>
</tr>
<tr>
<td>P018</td>
<td>2</td>
<td>21</td>
<td>167.6</td>
<td>67.3</td>
<td>22</td>
<td>240</td>
<td>1987.6</td>
<td>33.14</td>
<td>1613.5</td>
<td>24.78</td>
<td>1700.9</td>
<td>26.73</td>
<td>8</td>
</tr>
<tr>
<td>P019</td>
<td>2</td>
<td>21</td>
<td>153.7</td>
<td>54.5</td>
<td>29</td>
<td>300</td>
<td>2228.7</td>
<td>38.53</td>
<td>2257.8</td>
<td>39.18</td>
<td>2024.7</td>
<td>34.0</td>
<td>7</td>
</tr>
<tr>
<td>P020</td>
<td>2</td>
<td>19</td>
<td>163.8</td>
<td>58.2</td>
<td>22</td>
<td>90</td>
<td>2012.9</td>
<td>33.7</td>
<td>2061.6</td>
<td>34.79</td>
<td>2071.1</td>
<td>35.0</td>
<td>5</td>
</tr>
<tr>
<td>P021</td>
<td>2</td>
<td>19</td>
<td>177.2</td>
<td>80.5</td>
<td>22</td>
<td>180</td>
<td>2537.7</td>
<td>45.43</td>
<td>2424.5</td>
<td>42.49</td>
<td>2696.5</td>
<td>46.9</td>
<td>10</td>
</tr>
<tr>
<td>P022</td>
<td>2</td>
<td>21</td>
<td>156.1</td>
<td>65.9</td>
<td>22</td>
<td>280</td>
<td>2174.5</td>
<td>37.32</td>
<td>2143.0</td>
<td>36.61</td>
<td>2228.0</td>
<td>38.51</td>
<td>6</td>
</tr>
<tr>
<td>P023</td>
<td>2</td>
<td>20</td>
<td>160.0</td>
<td>59.1</td>
<td>22</td>
<td>180</td>
<td>1998.5</td>
<td>33.36</td>
<td>1613.5</td>
<td>24.78</td>
<td>1855.2</td>
<td>30.18</td>
<td>6</td>
</tr>
<tr>
<td>P024</td>
<td>2</td>
<td>18</td>
<td>167.6</td>
<td>66.4</td>
<td>22</td>
<td>210</td>
<td>2670.7</td>
<td>48.48</td>
<td>2611.6</td>
<td>47.09</td>
<td>2675.5</td>
<td>48.51</td>
<td>8</td>
</tr>
<tr>
<td>P025</td>
<td>2</td>
<td>21</td>
<td>166.4</td>
<td>49.1</td>
<td>22</td>
<td>210</td>
<td>2084.7</td>
<td>35.31</td>
<td>2025.9</td>
<td>34.0</td>
<td>2092.8</td>
<td>35.49</td>
<td>5</td>
</tr>
<tr>
<td>P026</td>
<td>1</td>
<td>19</td>
<td>184.8</td>
<td>76.4</td>
<td>22</td>
<td>720</td>
<td>2685.4</td>
<td>48.73</td>
<td>2772.8</td>
<td>50.69</td>
<td>2952.4</td>
<td>54.7</td>
<td>11</td>
</tr>
<tr>
<td>P027</td>
<td>1</td>
<td>18</td>
<td>193.0</td>
<td>78.2</td>
<td>27</td>
<td>330</td>
<td>2394.0</td>
<td>42.22</td>
<td>2418.9</td>
<td>42.78</td>
<td>2565.1</td>
<td>46.0</td>
<td>8</td>
</tr>
<tr>
<td>P028</td>
<td>1</td>
<td>20</td>
<td>1000</td>
<td>10</td>
<td>5</td>
<td>48.4</td>
<td>11</td>
<td>1</td>
<td>55.1</td>
<td>12</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P029</td>
<td>2</td>
<td>19</td>
<td>153.0</td>
<td>53.2</td>
<td>22</td>
<td>300</td>
<td>2178.1</td>
<td>37.4</td>
<td>2014.5</td>
<td>33.74</td>
<td>2201.5</td>
<td>37.92</td>
<td>5</td>
</tr>
<tr>
<td>P030</td>
<td>2</td>
<td>19</td>
<td>181.6</td>
<td>85.5</td>
<td>600</td>
<td>2690.9</td>
<td>48.86</td>
<td>11</td>
<td>5</td>
<td>55.1</td>
<td>12</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>P031</td>
<td>2</td>
<td>19</td>
<td>153.0</td>
<td>53.2</td>
<td>22</td>
<td>300</td>
<td>2178.1</td>
<td>37.4</td>
<td>2014.5</td>
<td>33.74</td>
<td>2201.5</td>
<td>37.92</td>
<td>5</td>
</tr>
<tr>
<td>sub</td>
<td>sex</td>
<td>age</td>
<td>height</td>
<td>weight</td>
<td>days</td>
<td>12MR1</td>
<td>12MR2</td>
<td>12MR3</td>
<td>MSR1</td>
<td>MSR2</td>
<td>MSR3</td>
<td>TR</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>--------</td>
<td>--------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>P032</td>
<td>2</td>
<td>19</td>
<td>157.5</td>
<td>66.4</td>
<td>27</td>
<td>420</td>
<td>2048.8</td>
<td>34.51</td>
<td>2090.4</td>
<td>35.44</td>
<td>2078.2</td>
<td>35.16</td>
<td>5</td>
</tr>
<tr>
<td>P033</td>
<td>2</td>
<td>19</td>
<td>162.6</td>
<td>66.1</td>
<td>22</td>
<td>1723.2</td>
<td>27.43</td>
<td>1830.7</td>
<td>29.63</td>
<td>1895.0</td>
<td>31.07</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>P034</td>
<td>1</td>
<td>19</td>
<td>179.1</td>
<td>101.8</td>
<td>22</td>
<td>2390.0</td>
<td>42.13</td>
<td>2515.0</td>
<td>44.93</td>
<td>2555.5</td>
<td>45.83</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>P035</td>
<td>1</td>
<td>21</td>
<td>179.1</td>
<td>72.7</td>
<td>22</td>
<td>800</td>
<td>2778.0</td>
<td>44.76</td>
<td>2907.4</td>
<td>47.84</td>
<td>2685.6</td>
<td>48.74</td>
<td>12</td>
</tr>
<tr>
<td>P036</td>
<td>1</td>
<td>28</td>
<td>179.2</td>
<td>70.5</td>
<td>22</td>
<td>720</td>
<td>2954.0</td>
<td>54.74</td>
<td>2907.4</td>
<td>55.11</td>
<td>2090.4</td>
<td>53.2</td>
<td></td>
</tr>
<tr>
<td>P037</td>
<td>2</td>
<td>24</td>
<td>164.1</td>
<td>55.5</td>
<td>12</td>
<td>630</td>
<td>2596.8</td>
<td>46.75</td>
<td>2443.5</td>
<td>43.33</td>
<td>2645.2</td>
<td>47.84</td>
<td>10</td>
</tr>
<tr>
<td>P038</td>
<td>2</td>
<td>22</td>
<td>175.0</td>
<td>72.5</td>
<td>17</td>
<td>420</td>
<td>2534.4</td>
<td>45.36</td>
<td>2390.0</td>
<td>44.93</td>
<td>2515.0</td>
<td>44.93</td>
<td>9</td>
</tr>
<tr>
<td>P039</td>
<td>2</td>
<td>24</td>
<td>165.8</td>
<td>68.5</td>
<td>36</td>
<td>630</td>
<td>2596.8</td>
<td>46.75</td>
<td>2443.5</td>
<td>43.33</td>
<td>2645.2</td>
<td>47.84</td>
<td>10</td>
</tr>
<tr>
<td>P040</td>
<td>2</td>
<td>22</td>
<td>160.8</td>
<td>56.0</td>
<td>24</td>
<td>400</td>
<td>2190.2</td>
<td>37.67</td>
<td>2217.1</td>
<td>41.06</td>
<td>2194.1</td>
<td>37.75</td>
<td>7</td>
</tr>
<tr>
<td>P041</td>
<td>1</td>
<td>22</td>
<td>179.1</td>
<td>113.0</td>
<td>22</td>
<td>0</td>
<td>2158.8</td>
<td>37.0</td>
<td>2104.0</td>
<td>35.74</td>
<td>2051.7</td>
<td>34.57</td>
<td>7</td>
</tr>
<tr>
<td>P042</td>
<td>1</td>
<td>25</td>
<td>170.1</td>
<td>60.0</td>
<td>28</td>
<td>150</td>
<td>3254.6</td>
<td>61.46</td>
<td>3314.0</td>
<td>62.79</td>
<td>3240.8</td>
<td>61.15</td>
<td>14</td>
</tr>
<tr>
<td>P043</td>
<td>1</td>
<td>23</td>
<td>166.5</td>
<td>67.5</td>
<td>18</td>
<td>180</td>
<td>2712.4</td>
<td>49.34</td>
<td>2672.6</td>
<td>48.45</td>
<td>2816.7</td>
<td>51.67</td>
<td>10</td>
</tr>
<tr>
<td>P044</td>
<td>2</td>
<td>22</td>
<td>170.5</td>
<td>60.0</td>
<td>15</td>
<td>300</td>
<td>2420.0</td>
<td>42.8</td>
<td>2372.5</td>
<td>41.74</td>
<td>2430.3</td>
<td>43.0</td>
<td>8</td>
</tr>
<tr>
<td>P045</td>
<td>1</td>
<td>23</td>
<td>176.0</td>
<td>78.0</td>
<td>13</td>
<td>300</td>
<td>2741.0</td>
<td>50.0</td>
<td>2822.0</td>
<td>51.79</td>
<td>2865.0</td>
<td>52.75</td>
<td>11</td>
</tr>
<tr>
<td>P046</td>
<td>2</td>
<td>22</td>
<td>165.0</td>
<td>72.0</td>
<td>15</td>
<td>240</td>
<td>2009.9</td>
<td>33.64</td>
<td>2015.0</td>
<td>33.75</td>
<td>1929.6</td>
<td>31.84</td>
<td>5</td>
</tr>
<tr>
<td>P047</td>
<td>1</td>
<td>23</td>
<td>172.0</td>
<td>70.5</td>
<td>16</td>
<td>40</td>
<td>2497.4</td>
<td>44.53</td>
<td>2534.4</td>
<td>45.36</td>
<td>2585.3</td>
<td>46.5</td>
<td>9</td>
</tr>
<tr>
<td>P048</td>
<td>2</td>
<td>23</td>
<td>158.4</td>
<td>60.5</td>
<td>16</td>
<td>63</td>
<td>1929.6</td>
<td>31.84</td>
<td>1965.1</td>
<td>32.64</td>
<td>2041.2</td>
<td>34.34</td>
<td>5</td>
</tr>
<tr>
<td>P049</td>
<td>2</td>
<td>24</td>
<td>167.0</td>
<td>58.0</td>
<td>17</td>
<td>33</td>
<td>2246.7</td>
<td>38.93</td>
<td>2342.1</td>
<td>41.06</td>
<td>2410.8</td>
<td>42.6</td>
<td>8</td>
</tr>
<tr>
<td>P050</td>
<td>1</td>
<td>22</td>
<td>199.1</td>
<td>79.0</td>
<td>22</td>
<td>0</td>
<td>2158.8</td>
<td>37.0</td>
<td>2104.0</td>
<td>35.74</td>
<td>2051.7</td>
<td>34.57</td>
<td>7</td>
</tr>
<tr>
<td>P051</td>
<td>1</td>
<td>21</td>
<td>167.4</td>
<td>54.0</td>
<td>12</td>
<td>1750.6</td>
<td>36.78</td>
<td>2216.6</td>
<td>38.26</td>
<td>2179.9</td>
<td>37.44</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>P052</td>
<td>1</td>
<td>28</td>
<td>179.3</td>
<td>68.0</td>
<td>18</td>
<td>2708.8</td>
<td>49.3</td>
<td>2832.0</td>
<td>52.0</td>
<td>2824.4</td>
<td>51.84</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>P053</td>
<td>1</td>
<td>19</td>
<td>173.3</td>
<td>70.5</td>
<td>16</td>
<td>40</td>
<td>2708.8</td>
<td>49.3</td>
<td>2534.4</td>
<td>45.36</td>
<td>2585.3</td>
<td>46.5</td>
<td>10</td>
</tr>
<tr>
<td>P054</td>
<td>2</td>
<td>30</td>
<td>164.2</td>
<td>56.5</td>
<td>25</td>
<td>480</td>
<td>2560.0</td>
<td>45.93</td>
<td>2569.6</td>
<td>46.15</td>
<td>2582.7</td>
<td>46.44</td>
<td>8</td>
</tr>
<tr>
<td>P055</td>
<td>1</td>
<td>30</td>
<td>175.8</td>
<td>76.0</td>
<td>43</td>
<td>350</td>
<td>3115.2</td>
<td>58.34</td>
<td>3225.1</td>
<td>60.8</td>
<td>3294.7</td>
<td>62.35</td>
<td>11</td>
</tr>
<tr>
<td>P056</td>
<td>1</td>
<td>26</td>
<td>173.7</td>
<td>79.5</td>
<td>36</td>
<td>50</td>
<td>3083.2</td>
<td>57.63</td>
<td>3154.9</td>
<td>59.23</td>
<td>3037.4</td>
<td>56.6</td>
<td>11</td>
</tr>
<tr>
<td>P057</td>
<td>1</td>
<td>23</td>
<td>177.6</td>
<td>67.5</td>
<td>24</td>
<td>240</td>
<td>3410.0</td>
<td>64.93</td>
<td>3083.2</td>
<td>60.76</td>
<td>3037.4</td>
<td>56.6</td>
<td>12</td>
</tr>
<tr>
<td>P058</td>
<td>2</td>
<td>20</td>
<td>166.1</td>
<td>67.0</td>
<td>26</td>
<td>105</td>
<td>2155.4</td>
<td>36.89</td>
<td>2212.0</td>
<td>38.35</td>
<td>2156.8</td>
<td>36.92</td>
<td>6</td>
</tr>
<tr>
<td>P059</td>
<td>1</td>
<td>28</td>
<td>175.1</td>
<td>78.0</td>
<td>13</td>
<td>600</td>
<td>2877.0</td>
<td>53.0</td>
<td>2932.0</td>
<td>54.25</td>
<td>2913.7</td>
<td>53.84</td>
<td>10</td>
</tr>
<tr>
<td>P060</td>
<td>2</td>
<td>32</td>
<td>162.3</td>
<td>50.5</td>
<td>22</td>
<td>516</td>
<td>3083.2</td>
<td>57.63</td>
<td>3154.9</td>
<td>59.23</td>
<td>3037.4</td>
<td>56.6</td>
<td>11</td>
</tr>
<tr>
<td>P061</td>
<td>1</td>
<td>28</td>
<td>164.6</td>
<td>66.5</td>
<td>17</td>
<td>2822.5</td>
<td>51.8</td>
<td>2838.4</td>
<td>52.15</td>
<td>2876.0</td>
<td>53.0</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>P062</td>
<td>2</td>
<td>33</td>
<td>164.8</td>
<td>58.0</td>
<td>22</td>
<td>900</td>
<td>3028.0</td>
<td>56.39</td>
<td>3036.1</td>
<td>56.57</td>
<td>3001.2</td>
<td>55.79</td>
<td>10</td>
</tr>
<tr>
<td>sub</td>
<td>sex</td>
<td>age</td>
<td>height</td>
<td>weight</td>
<td>days</td>
<td>pamin</td>
<td>12MR1</td>
<td>12MR2</td>
<td>12MR3</td>
<td>MSR1</td>
<td>MSR2</td>
<td>MSR3</td>
<td>TR</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>--------</td>
<td>--------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>meters</td>
<td>VO2</td>
<td>meters</td>
<td>VO2</td>
<td>stage</td>
<td>shut</td>
<td>VO2</td>
</tr>
<tr>
<td>P063</td>
<td>1</td>
<td>20</td>
<td>173.0</td>
<td>83.1</td>
<td>36</td>
<td>300</td>
<td>2777.1</td>
<td>50.78</td>
<td>2799.0</td>
<td>51.27</td>
<td>2725.2</td>
<td>49.62</td>
<td>10</td>
</tr>
<tr>
<td>P064</td>
<td>1</td>
<td>21</td>
<td>188.9</td>
<td>90.8</td>
<td>22</td>
<td>500</td>
<td>3053.7</td>
<td>57.0</td>
<td>3169.0</td>
<td>59.54</td>
<td>3079.6</td>
<td>57.55</td>
<td>13</td>
</tr>
<tr>
<td>P065</td>
<td>1</td>
<td>22</td>
<td>178.6</td>
<td>86.3</td>
<td>36</td>
<td>180</td>
<td>2142.0</td>
<td>36.59</td>
<td>2240.5</td>
<td>38.79</td>
<td>2273.9</td>
<td>39.54</td>
<td>7</td>
</tr>
<tr>
<td>P066</td>
<td>1</td>
<td>19</td>
<td>179.7</td>
<td>88.6</td>
<td>29</td>
<td>800</td>
<td>3012.7</td>
<td>56.05</td>
<td>3000.0</td>
<td>55.77</td>
<td>2998.7</td>
<td>55.74</td>
<td>11</td>
</tr>
<tr>
<td>P067</td>
<td>1</td>
<td>19</td>
<td>184.1</td>
<td>98.5</td>
<td>22</td>
<td>250</td>
<td>2799.0</td>
<td>51.27</td>
<td>2910.0</td>
<td>53.76</td>
<td>2920.5</td>
<td>54.0</td>
<td>11</td>
</tr>
<tr>
<td>P068</td>
<td>1</td>
<td>20</td>
<td>180.9</td>
<td>80.4</td>
<td>22</td>
<td>300</td>
<td>2875.8</td>
<td>53.0</td>
<td>2913.0</td>
<td>53.82</td>
<td>2768.4</td>
<td>50.59</td>
<td>10</td>
</tr>
<tr>
<td>P069</td>
<td>1</td>
<td>20</td>
<td>172.2</td>
<td>62.2</td>
<td>17</td>
<td>60</td>
<td>2122.5</td>
<td>36.15</td>
<td>2281.1</td>
<td>39.7</td>
<td>2328.8</td>
<td>40.76</td>
<td>9</td>
</tr>
<tr>
<td>P070</td>
<td>1</td>
<td>21</td>
<td>178.7</td>
<td>78.9</td>
<td>17</td>
<td>120</td>
<td>2143.0</td>
<td>36.61</td>
<td>2291.4</td>
<td>39.93</td>
<td>2336.9</td>
<td>40.95</td>
<td>7</td>
</tr>
</tbody>
</table>

- completed 12MR, MSR, TR
- completed 12MR, MSR
- did not complete study

1 = male
2 = female
Variance Components Estimation Procedure

Class Level Information

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub</td>
<td>60</td>
<td>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 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60</td>
</tr>
<tr>
<td>instrument</td>
<td>2</td>
<td>1 2</td>
</tr>
<tr>
<td>day</td>
<td>3</td>
<td>1 2 3</td>
</tr>
</tbody>
</table>

Number of Observations Read 360
Number of Observations Used 360

Dependent Variable: score

Type 1 Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub</td>
<td>59</td>
<td>24766</td>
<td>419.763053</td>
</tr>
<tr>
<td>instrument</td>
<td>1</td>
<td>0.298714</td>
<td>0.298714</td>
</tr>
<tr>
<td>day</td>
<td>2</td>
<td>28.164522</td>
<td>14.082261</td>
</tr>
<tr>
<td>sub*instrument</td>
<td>59</td>
<td>173.205570</td>
<td>2.935688</td>
</tr>
<tr>
<td>sub*day</td>
<td>118</td>
<td>1129.478112</td>
<td>9.571848</td>
</tr>
<tr>
<td>instrument*day</td>
<td>2</td>
<td>0.597427</td>
<td>0.298714</td>
</tr>
<tr>
<td>Error</td>
<td>118</td>
<td>346.411139</td>
<td>2.935688</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>359</td>
<td>26444</td>
<td></td>
</tr>
</tbody>
</table>

Type 1 Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub</td>
<td>Var(Error) + 2 Var(sub<em>day) + 3 Var(sub</em>instrument) + 6 Var(sub)</td>
</tr>
<tr>
<td>instrument</td>
<td>Var(Error) + 60 Var(instrument<em>day) + 3 Var(sub</em>instrument) + 180 Var(instrument)</td>
</tr>
<tr>
<td>day</td>
<td>Var(Error) + 60 Var(instrument<em>day) + 2 Var(sub</em>day) + 120 Var(day)</td>
</tr>
<tr>
<td>sub*instrument</td>
<td>Var(Error) + 3 Var(sub*instrument)</td>
</tr>
<tr>
<td>sub*day</td>
<td>Var(Error) + 2 Var(sub*day)</td>
</tr>
<tr>
<td>instrument*day</td>
<td>Var(Error) + 60 Var(instrument*day)</td>
</tr>
<tr>
<td>Error</td>
<td>Var(Error)</td>
</tr>
<tr>
<td>Corrected Total</td>
<td></td>
</tr>
</tbody>
</table>
Variance Components Estimation Procedure

Type 1 Estimates

<table>
<thead>
<tr>
<th>Variance Component</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var(sub)</td>
<td>68.36520</td>
</tr>
<tr>
<td>Var(instrument)</td>
<td>5.0241E-15</td>
</tr>
<tr>
<td>Var(day)</td>
<td>0.05956</td>
</tr>
<tr>
<td>Var(sub*instrument)</td>
<td>-2.53E-13</td>
</tr>
<tr>
<td>Var(sub*day)</td>
<td>3.31808</td>
</tr>
<tr>
<td>Var(instrument*day)</td>
<td>-0.04395</td>
</tr>
<tr>
<td>Var(Error)</td>
<td>2.93569</td>
</tr>
</tbody>
</table>
The GLM Procedure

Class Level Information

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub</td>
<td>60</td>
<td>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 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60</td>
</tr>
<tr>
<td>instrument</td>
<td>2</td>
<td>1 2</td>
</tr>
<tr>
<td>day</td>
<td>3</td>
<td>1 2 3</td>
</tr>
</tbody>
</table>

Number of Observations Read 360
Number of Observations Used 360
The GLM Procedure

Dependent Variable: score

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>241</td>
<td>26097.76450</td>
<td>108.28948</td>
<td>36.89</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>118</td>
<td>346.41114</td>
<td>2.93569</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>359</td>
<td>26444.17564</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square  Coeff Var  Root MSE  score Mean
0.986900  3.953699  1.713385  43.33625

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub</td>
<td>59</td>
<td>24766.02016</td>
<td>419.76305</td>
<td>142.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>instrument</td>
<td>1</td>
<td>0.29871</td>
<td>0.29871</td>
<td>0.10</td>
<td>0.7503</td>
</tr>
<tr>
<td>day</td>
<td>2</td>
<td>28.16452</td>
<td>14.08226</td>
<td>4.80</td>
<td>0.0099</td>
</tr>
<tr>
<td>sub*instrument</td>
<td>59</td>
<td>173.20557</td>
<td>2.93569</td>
<td>1.00</td>
<td>0.4900</td>
</tr>
<tr>
<td>sub*day</td>
<td>118</td>
<td>1129.47811</td>
<td>9.57185</td>
<td>3.26</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>instrument*day</td>
<td>2</td>
<td>0.59743</td>
<td>0.29871</td>
<td>0.10</td>
<td>0.9033</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub</td>
<td>59</td>
<td>24766.02015</td>
<td>419.76305</td>
<td>142.99</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>instrument</td>
<td>1</td>
<td>0.29871</td>
<td>0.29871</td>
<td>0.10</td>
<td>0.7503</td>
</tr>
<tr>
<td>day</td>
<td>2</td>
<td>28.16452</td>
<td>14.08226</td>
<td>4.80</td>
<td>0.0099</td>
</tr>
<tr>
<td>sub*instrument</td>
<td>59</td>
<td>173.20557</td>
<td>2.93569</td>
<td>1.00</td>
<td>0.4900</td>
</tr>
<tr>
<td>sub*day</td>
<td>118</td>
<td>1129.47811</td>
<td>9.57185</td>
<td>3.26</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>instrument*day</td>
<td>2</td>
<td>0.59743</td>
<td>0.29871</td>
<td>0.10</td>
<td>0.9033</td>
</tr>
</tbody>
</table>
options ps = 55;
options ls = 79;
pageno = 1;
data run;
input
  sub i1d1 i1d2 i1d3 i2d1 i2d2 i2d3;
instrument=1;day=1;score=i1d1;output;
instrument=1;day=2;score=i1d2;output;
instrument=1;day=3;score=i1d3;output;
instrument=2;day=1;score=i1d1;output;
instrument=2;day=2;score=i1d2;output;
instrument=2;day=3;score=i2d3;output;
cards;
1  44.34 43.88 46.71 46.10 49.60 52.50
2  42.36 44.80 41.02 39.20 40.20 38.20
3  44.25 43.52 44.27 39.20 40.50 40.20
4  41.41 33.42 34.09 44.50 43.00 41.10
5  38.79 39.16 35.59 35.70 36.40 37.10
6  28.38 27.74 25.91 27.20 29.10 28.70
7  41.40 39.44 43.17 39.90 40.50 43.00
8  34.60 43.08 42.17 47.10 44.50 39.60
9  42.61 42.24 35.59 41.80 44.50 43.60
10  45.80 45.66 45.97 46.10 49.30 49.90
11  44.61 46.21 47.24 41.50 46.60 47.70
12  29.26 28.61 31.72 27.60 29.10 28.00
13  33.38 24.78 26.73 36.80 33.30 34.30
14  44.52 45.74 46.72 42.70 46.10 47.10
15  33.14 24.78 26.73 40.20 37.10 37.10
16  38.53 39.18 33.97 37.50 40.20 41.50
17  33.70 34.79 35.00 30.20 27.60 31.00
18  45.43 42.90 48.99 48.70 51.70 51.70
19  37.32 36.61 38.51 36.40 39.60 37.10
20  33.38 24.78 30.18 33.30 35.40 33.30
21  48.48 47.09 48.51 42.70 44.90 46.80
22  35.31 33.99 35.49 32.20 35.00 32.20
23  48.73 50.69 54.70 51.10 51.40 52.80
24  42.22 42.78 46.05 40.20 43.00 43.30
25  37.40 33.74 37.92 32.90 36.80 37.50
26  34.51 35.44 35.16 31.00 32.20 32.90
27  27.43 29.63 31.07 30.20 27.20 33.60
28  42.13 44.93 45.83 44.90 46.80 47.70
29  50.80 44.76 48.74 54.30 47.70 53.40
30  48.67 49.93 47.60 60.60 44.20 60.60
31  54.74 53.63 55.11 46.10 46.10 47.10
32  45.36 43.33 45.18 42.70 42.10 43.60
33  46.75 43.33 47.84 48.40 47.70 50.50
34  37.67 38.27 37.75 37.10 34.70 36.80
35  36.96 35.74 34.57 38.20 37.10 37.50
36  61.46 62.79 61.15 60.90 60.90 61.10
37  49.34 48.45 51.67 48.40 49.30 51.70
38  42.80 41.74 43.03 41.80 43.90 41.80
39  49.98 51.79 52.75 50.80 53.70 54.30
40  33.64 33.75 31.84 31.40 32.20 33.60
41  44.53 45.36 46.50 46.80 48.70 44.50
Proc varcomp method=type1;
  class sub instrument day;
  model score=sub instrument day
    sub*instrument
    sub*day
    instrument*day;
Proc glm;
  class sub instrument day;
  model score=sub instrument day
    sub*instrument
    sub*day
    instrument*day;
run;

APPENDIX D

Data collection journal
4.8.08 – Trial day. Met with exercise physiology students to set up schedule. Marked course on fieldhouse floor. Did an abbreviated shuttle run trial to make sure equipment worked.

4.10.08 – First day with participants. Participants signed informed consent and filled out health questionnaire. No participants were initially excluded. One participant did not complete the first testing session (flu-like symptoms). I conducted the shuttle run testing, with Lance and Cassie conducting the 12-minute run. Lance informs me that some students run together during the duration of the 12-minute run, and that he is working to discourage that behavior. Weather is cloudy and cool, but no rain.

4.15.08 – Second day with participants. Two new participants showed up today. Weather is a little warmer than Tuesday, with some sun.

4.17.08 – Third day with participants. Two participants who were conducting the shuttle run today noted distinct factors that may have affected their performance (knee pain, asthma). Some participants are beginning to voice mild displeasure at repeating the same test three times in a row. Weather cooler, with some sun.

4.22.08 – Fourth day with participants. A few participants missed today; I e-mailed them to see if they wanted to stay in the study. Weather outside was cold with light drizzle – probably the “worst” weather we’ve had so far, but actually not too bad. Winds were calm.

4.24.08 – Fifth day with participants. Most (if not all) students had a big assignment due today, so there was some missed sleep all around. Participants still seemed to give maximal effort, however. Cassie assisted me with data collection in on the 20-meter shuttle run (1st time), while Lance and Nelson took care of the 12-minute run. Weather was cool (40s), but no precipitation and calm winds.

5.1.08 – Sixth day with participants. Fewer students showed up today – I’m assuming that’s because they know they can miss one or two sessions and still make the four-week window. Cassie and Lance collected data on the 12-minute run – actually had some breaks in the clouds today and saw some sun. Still cool (40s) and calm winds. 18 individuals completed the field tests today – I will e-mail them to see if they are interested in the treadmill portion of the test. I will also personally e-mail each of the remaining 12 Linfield participants to encourage them to finish the study next week.

5.6.08 – Seventh day with Linfield participants. Picking up the stragglers now – both the shuttle run sessions today had less than 3 participants. Cassie and Lance collected data on 5 12-minute run trials, while Nelson helped me with the shuttle run. Weather was cool (low 40s), cloudy, but no rain.

5.7.08 – First day testing with OSU students. Collected data at Linus Pauling Middle School track. Weather was in the 40s with a light wind. Participants show good enthusiasm.
5.8.08 – Treadmill testing at Linfield College. Four participants, all met two or more criteria for a true max test. Janet Peterson helped with all four tests.

5.9.08 – Second day of testing at OSU (Linus Pauling). 40s again, but with less wind. Two participants reported feeling fatigued from the week before testing, and subsequently scored lower on the run.

5.13.08 – Final day of field testing at Linfield College. One participant completed the shuttle run, and appeared to be in good spirits. This participant completed the study late because he was hit in the knee with a softball during the earlier portion of the study, and felt that injury would affect his performance at that time.

5.14.08 – OSU testing; first day in WB gym using the shuttle run. Minor miscommunication with several of the participants regarding location, but they did make it in time to test. Participants were divided as to whether this test felt harder/easier: those that felt it was easier liked the “pacing” aspect of the shuttle run, while the other individual (P040) felt the test was harder for the same reason.

5.15.08 – Max testing at Linfield College. All participants met at least 2 of the 3 criteria for a max test. All participants were also very highly motivated to do this sort of test, based upon prior experience with a similar test, or interest in endurance training.

5.16.08 – Shuttle run testing at OSU, 3 participants. Participants report general fatigue from the week.

5.19.08 – Tested one participant (P036) at OSU, shuttle run. Quite warm in the gymnasium where the testing took place.

5.21.08 – Shuttle run testing at OSU, 3 participants. Much cooler than 5/19. Participants are glad to be finished with the shuttle run.

5.23.08 – 12-minute run testing for OSU participants at Linus Pauling. Three new participants started today, one student completed the study. Weather was quite cool (50 degrees or so) and cloudy. Calm to no winds.

5.28.08 – 12-minute run testing at Linus Pauling. Several new participants today. Two groups for run. Moderate winds, drizzle for last group of participants (2 groups). One participant noted that the wind on the backstretch may have played a role in his performance. I think this may have been more a psychological factor, since standing at the start line, the wind was not highly noticeable.
5.30.08 – Shuttle run testing at OSU. Two groups of participants (one new); four participants ran in each trial.

Ideas: - Females do better (relatively) on shuttle run because of lower COG? - differences between males and females not taken in account when setting table for estimating VO₂ from shuttles (from Chun et al. 2000; see Ramsbottom et al. 1988)
- Split trial comparison into two groups (top 50% and lower 50%) due to differential effect of fatigue on the two groups? - weaker correlation between PACER and 12-minute run in trained individuals (Chun et al. 2000) … perhaps trained individuals can “cheat” a particular test? Must compare to absolute treadmill test when possible.

6.2.08 – Shuttle run testing at OSU. Started a little earlier this morning than usual, but the participants did not seem to mind this early start. One group of 5 participants today.

6.4.08 – 12-minute run testing at Linus Pauling. Participants mentioned that it felt abnormally cold this morning (lower 40s?) – participants noted this prior to their testing. Light winds.

6.6.08 – 12-minute run testing at Linus Pauling track. Warmer than 6/4 with sun. Participants were in good spirits.

6.9.08 – Shuttle run testing at OSU. Individuals showed good motivation this morning. Three participants in trial 1, four participants in trial 2.

Ideas: - Effect of missing one day with individual of “close” fitness?
- P045 has “learned” to jog just easily enough during early stages to make the beeps on each shuttle. This appears to allow him to keep more “in the tank” for the later stages.

6.10.08 – Shuttle run testing, one participant (P047) @ OSU. This participant had to complete the test today so that he would get a full set of data before leaving for the summer.

6.11.08 – Linus Pauling (8AM). 4 participants. Good effort by all participants.

6.11.08 – Shuttle run testing @ OSU (9AM). 3 Participants. Floor seemed a little dusty this morning, although perhaps that is similar to the norm.

6.12.08 – Shuttle run testing at OSU. New participant was quite tall – had some trouble with the change of direction involved with the test.

6.13.08 – 12-minute run testing at Linus Pauling (8AM). One participant – several no-shows this morning. Cool and bright sun.
6.13.08 – Maximal treadmill test this morning (10AM). Although participant appeared to give a strong effort, the participant only reached 1 of 3 criteria for maximal test. Participant was asked to return for a second maximal test.

6.13.08 – 12-minute run testing at Linus Pauling (1PM). Working with no-shows from this morning – had to test at the later time period because participants were leaving town in the near future. Somewhat warm (80s) and bright sun.

6.18.08 – Second maximal treadmill test for participant from 6/13. Did not take participant up to the same speed today, but worked more with an increase in incline. Participant reached 2 of 3 criteria for max; minute-average oxygen uptake was within 0.6 mL/kg/min of data from 6/13.

6.19.08 – Shuttle run testing @ OSU. Two good efforts this morning.

6.23.08 – Shuttle run testing at OSU. Only one show this morning. Cooler in gymnasium this morning.

6.23.08 – Four maximal treadmill tests this morning (9AM-12PM). Found a small hole in one of the hoses this morning – made a note of this and removed tube from circulation. All four participants reached 2 of the 3 criteria required for a maximal test.

6.25.08 – Last 12-minute run for a single participant this morning (8 AM) at Linus Pauling. Track was a very busy place this morning, although it did not directly affect the participant’s progress. Bright sun.

7.1.08 – Shuttle run test at OSU for three new participants this morning. Good effort by all three participants.

7.3.08 – Testing at Linus Pauling. Cooler this morning, bright sun. Had the track to ourselves (4 participants). First track trial for all four participants.

7.6.08 – Testing at Linus Pauling. Pretty cool (low 50s) and overcast. Mild breeze – participants made note of it, but said that they did not feel that it was a major determinant in their performances.

7.8.08 – Shuttle run testing at OSU. Two participants (several no-shows today). Good effort by both participants.

7.9.08 – Shuttle run testing (OSU). Only one participant – several no-shows again. Participant indicated that even though the shuttle run was shorter, she preferred the 12-minute run because it was a movement pattern that she was “used to.”

7.10.08 – Three participants, shuttle run testing at OSU. Good effort by all – warm this morning in the gym.
7.11.08 – Shuttle run testing at OSU – three participants. Two new participants, which is always nice. Participants cheered each other on this morning.

7.12.08 – 12-minute run testing at Linus Pauling Middle School. Three participants. Tested at 8AM, but still quite warm and bright sun. Strong effort by all three individuals.

7.14.08 – One treadmill test this morning – participant reached max as shown by heart rate and plateau. Probably should have taken her up another 0.5 mph, as she was able to complete 14+ minutes of testing. Also tested two individuals following in the shuttle run. Both gave strong efforts. One participant (this was her second shuttle run) mentioned that she was better able to “time” her shuttles because of (1) her experience, she felt more used to the timing, and (2) having someone else to run with also helped her time her shuttles (helped her gauge her speed).

7.15.08 – Tested two individuals at Linus Pauling. One participant was new, and seemed to have trouble pacing herself. She did, however, complete the test in good spirits. Cooler today.

7.17.08 – Tested one individual in the 12-minute run at Linus Pauling. Several no-shows this morning. Still cooler and sunny (60s?).

7.18.08 – Tested one individual this morning at Linus Pauling. Several people did not show up for the test again this morning – perhaps due to the last week of summer courses at OSU. Participant did appear to give a good effort.

7.20.08 – One individual testing at Linus Pauling at 10AM. Sunny and warmer than previous days – probably in the mid 70s. Light wind. Participant was in good spirits. Also tested two individuals in the shuttle run at OSU at 11:30AM. Participants in the shuttle run noted that they “knew” when they had missed, and that matched up with the calls I made as the test administrator.

7.23.08 – Tested one individual at 8AM at Linus Pauling. No other individuals were on the track. Cooler today, with clear skies. (7 PM) Tested a second individual today in Portland on the Duniway track. Many individuals present on track during test, but participant appeared to have a clear path at all times in lane 1. Sunny and breezy – quite cool in the shade. Track was a yard track – required conversion to meters for laps completed.

7.24.08 – Tested two new participants in the shuttle run at 8AM. Both seem to be quite energetic about the study, as they are members of the OSU triathlon club. Good effort by both participants.

7.26.08 – Tested two individuals (triathletes) at Linus Pauling at 8AM. Sunny and clear.
7.29.08 – Tested four individuals today in the 12-minute run. One individual tested at 7:30AM, the other three (one new!) at 8AM. There were other individuals on the track running throughout all tests. There was some light drizzle before the test, but it was overcast with calm winds for both tests.

7.30-8.21.08 – Did not make daily records. Made notes on sheets when special occasion was needed.

9.11.08 – Recruited individuals from the fall term nutrition class at Linfield College. Asked only males to volunteer so that it would bring the final numbers close to 50/50. Participants were offered extra credit for participation in the study.

9.16.08 – Tested eight individuals at Linfield College. The individuals were tested in three groups – 7:30AM (2 individuals), 8AM (4 individuals), and 9AM (2 individuals). All seemed in good spirits and were eager to begin participation in the study.

9.18.08 – Tested seven individuals at Linfield College. Forgot measuring wheel/shuttle run test apparatus at home, but was able to borrow a measuring wheel at the test site. Two of the three groups were scheduled to run today, so it only resulted in a change for one group (the last, at 9AM). Much cooler today – overcast and in the 60s; perhaps even some light drizzle. Participants seem in good spirits.

9.23.08 – Second day with the second Linfield cohort. Two fieldhouse testing sessions (shuttle run) and one outdoor session (12-minute run). The second fieldhouse session was split in two because one individual was late to the session, but he still completed the test session in good spirits. In the third session, lane 1 was obstructed from crowd control equipment from a football game, so participants stayed in lane 2 throughout the test. Bright sun and cool (~50 degrees).

9.25.08 – Tested all individuals on the track, 12-minute run (Linfield College). Relatively warm (low 50s) and some sun, with little wind. Participants were in good spirits.

9.30.08 – Tested all individuals on the shuttle run in the Linfield College fieldhouse. Individuals seemed more comfortable with the test, as it was repeated for most participants. I’m beginning to suspect that a difference in variability may exist between very fit and less fit individuals.

10.2.08 – Testing at Linfield. Two shuttle runs at 8AM, two shuttle runs at 9AM. Participants seem comfortable with the testing protocol.

10.7.08 – Four shuttle runs today – one at 7:30AM, the other three at 8:00AM. Participants in the second group were particularly competitive today – should bring up the effect of competition on fitness evaluation the discussion of the dissertation.

10.9.08 – Participants did not show for testing today (Linfield College).
10.14.08 – Slim showings today (Linfield College). One individual showed for a 12-minute run at 8AM, another for a shuttle run at 8:30AM. Good weather for 12-minute run – no rain and crisp (40s).

10.16.08 – Participants did not show for testing again today (Linfield College). Was able to gather descriptive data (height, weight) from one individual.

10.21.08 – Last day of testing at Linfield. Cold and foggy for 7:45 12-minute run, followed by a fieldhouse shuttle run and a treadmill max test. Participants seem glad to be finished.
APPENDIX E

Proposal document
INTRODUCTION

The physiological parameter most closely linked to human endurance performance is maximal aerobic power (Yoshida et al., 1993). As such, accurate tests of this variable are very important, as the data are critical to prescribing physical activity in athletes of all ages. Aerobic power is most accurately measured in a laboratory setting by collecting expired gases (Thoden, 1990). However, this technique requires sophisticated equipment and test administrators specifically trained in the operation of this equipment. Access to such resources is not always available or feasible, and a real need exists for accurate tests that can be used by a variety of test administrators in a non-laboratory environment.

As a result of this demand, researchers have developed a number of tests that can be used outside the laboratory to estimate maximal aerobic power. These “field” tests are generally more variable than their laboratory counterparts, as they (1) estimate aerobic power from the performance of a physical task, rather than measuring oxygen uptake, and (2) are conducted in an environment that is less controlled than the laboratory, under differing conditions that may affect an individual’s performance. High variability reduces the utility of these field tests, as a test administrator cannot be sure a test score under such conditions is indicative of the true aerobic power score for that individual. Although field tests have shown strong relationships with a criterion measure in active populations ($r > 0.85$) (Cooper, 1968; Aziz et al., 2005), correlations associated with commonly used field tests such as Cooper’s 12-minute Run (TMR) or the Multistage Shuttle Run (MSR) in such populations are typically moderate ($r < 0.70$) (Wanamaker, 1970; O’Gorman et al., 2000). This reduction in test validity
underscores the necessity for a test administrator to understand and address the factors specific to a particular field test that may act to increase score variability.

Most studies that have investigated the variability associated with field tests of aerobic power have used Pearson’s product-moment correlation coefficient (PPM) to quantify the magnitude of test-retest reliability associated with these tests. The TMR and MSR have demonstrated good reliability, as test-retest r-values for these two tests have been reported as high as 0.98 (Krustrup et al., 2003) for the MSR and 0.94 (Doolittle and Bigbee, 1968) for the TMR. However, recent research suggests that the Pearson Product-Moment underestimates the variability associated with such tests. A 95% limits of agreement (LoA) analysis by Fairbrother et al. (2005) suggested that a score of 45 mL/kg/min on the MSR could be interpreted as a value as low as 40 mL/kg/min, or as high as 49 mL/kg/min, a characteristic that would render the test useless for detecting small changes in aerobic power. Moreover, it appears that a simple test-retest analysis may not give a complete picture of the variability associated with the MSR. In a method similar to that used by Fairbrother et al. (2005), Lamb et al. (2007) showed that variability is higher between the first and second administrations of the MSR than the second and third administrations of the test. As a result of these findings, it appears that researchers should apply a method other than the PPM when examining variability within tests such as the TMR or MSR.

The LoA technique used by Fairbrother et al. (2005) and Lamb et al. (2007) was first suggested by Bland and Altman (1986) for use in large heterogeneous samples where score variability in the independent variable may differ over the range of the dependent variable. Analysis via LoA better captures this variability than does
the PPM, but the method of visual inspection intrinsic to the LoA technique limits researchers to the investigation of one source of variability at a time. Moreover, LoA analysis does not allow researchers to consider the contribution that an interaction between two (or more) individual sources of variability may have on the total variability of a data set. This last matter may be of particular importance in measures of human performance, where multiple variables interact to produce a test score. Thus, an ideal measure quantifying variability within field tests of aerobic power should be sensitive to (1) variability over the range of the independent variable, and (2) the multifaceted relationships that can arise from tests that use human participants.

A statistical tool that fulfills these two requirements is generalizability theory, or G-theory. While standard analyses partition total variance into treatment variance and error variance, G-theory uses multiple repetitions of a treatment under selected conditions to describe the amount of error variance attributable to each selected condition (a facet). Common facets include test occasion and test administrator, but a G-study can be designed to examine any facet a researcher might deem important for investigation.

G-theory is especially popular in psychometric studies (Fraley and Kreuger, 2007), but has also shown promise in the field of human performance, where it has been used to examine tests of body composition (Jackson et al., 1988), isometric force (Roebroeck et al., 1993), and joint range of motion (Nadeau et al., 2007). Since G-studies can describe the major contributors to error variance within a test, it also allows researchers to make specific recommendations by which a particular test protocol may be improved. For example, when Roebroeck et al. (1993) found that
participants showed large isometric force variability when measured by different therapists on different test occasions, they were able to conclude that this variability would be best reduced by adding an initial “familiarization” session. Although no attempt has yet been made to apply the G-study technique to measures of aerobic power, this lack of application does not preclude the potential utility of the G-theory in developing measures of aerobic power, as the work by Lamb (2007) evidences.

A routine follow-up analysis to a G-study is a decision study, or D-study. A D-study uses the information gathered as part of the G-study, and mathematically models the changes in error variability following an increase or decrease in the quantity of a facet variable. As a result, researchers can choose the “optimal” magnitude of a facet that minimizes error variability while also conserving the available research-related resources. Roebroeck et al. (1993) illustrated this point during their analysis of isometric force variability, showing that in order to maximize reliability on a single test occasion by a single therapist, it was best to employ at least three test repetitions. Again, D-study has not been used to examine field tests of aerobic power, but such an analysis would be useful in describing how to use existing field tests to best measure aerobic power.

This study will utilize G-theory to determine the relative contribution of individual, test, and test occasion (as well as the interaction among these variables) to the total variability associated with the TMR and MSR, two of the most frequently used field tests of aerobic power (Tomkinson and Olds, 2007). These results will be used to make recommendations as to how test administrators may reduce the variability associated with each field test. In addition, these data will be combined
with data from a laboratory criterion test of aerobic power, and will permit researchers
to examine the criterion-related validity of the two field tests. Establishing the validity
of the two field tests in such a manner is especially important to the generalizability of
the study findings, as a validity coefficient similar in magnitude to that found in prior
studies would suggest the tests were administered using similar protocols.
RESEARCH AIMS

1. To investigate the largest sources of variability in Leger’s Multistage 20-meter Shuttle Run and Cooper’s 12-minute Run. Possible important sources of variability could include (a) instrument-to-instrument variability, (b) day-to-day variability, (c) individual-test interaction variability, and (d) day-test interaction variability.

2. To suggest means by which the variability associated with the selected tests might be reduced. Possible modifications include the (a) inclusion of a “habituation” trial to accustom participants to a particular test, (b) subtraction of test occasions to conserve resources and/or mitigate a possible training/fatigue effect, and (c) mitigation of important environmental factors during the administration of a particular test.

3. To determine the validity of Leger’s Multistage 20-meter Shuttle Run and Cooper’s 12-minute Run in the measured population.
MATERIALS AND METHODS

PARTICIPANTS

Participants in this study will be individuals between 18 and 35 years old, and be recruited from both the Corvallis and Greater Portland (Oregon) areas. Approximately 60 participants will be recruited for this study. Approximately the same number of males and females will be included in the study to increase the generalizability of the findings across the two sexes.

INSTRUMENTS

Each individual will complete three trials of two commonly used field tests of aerobic power. These tests will include Cooper’s 12-minute Run (Cooper, 1968) and Leger’s Multistage 20-meter Shuttle Run (Leger et al., 1984). These two field tests were selected in order to compare the sources of variability among the different types of tests regularly administered in field settings (Tomkinson and Olds, 2007). In addition to these two tests, a sub-sample of study participants will volunteer to complete a maximal treadmill VO\textsubscript{2max} test. Due to time constraints, only one-third of the total participant pool will be tested using the maximal treadmill test. The data resulting from this supplemental test will be important to directly measuring the criterion-related validity of the field tests, as well as indirectly assessing the construct-related validity of the field tests when no criterion measure is available.
PROCEDURES

General considerations. Participants will be asked to maintain their current activity level and refrain from any vigorous physical activity for the 24-hour period prior to a testing session. Participants will be placed into non-random groups of three to five, with test groups consisting of participants with similar schedules that will permit simultaneous testing. Groups will conduct the field tests in an order determined by a web-based random number generator (www.random.org, Mads Haahr). Each day that the participants are tested will be followed by at least one non-test day to minimize the effects of fatigue on the test results. Participants must complete each field test as a group, with a minimum of three participants present during a singular test administration. In the event that an individual does not complete a test with his/her cohort, that individual will continue the test battery with his/her cohort, and complete the missing test at the end of the test battery with another cohort. Participants must complete all six field tests and the treadmill test (if appropriate) within a four-week test period to minimize any training effect that may be present.

Prior to the initial test, individuals will complete a short health history questionnaire. In the event that an individual demonstrates more than one risk factor for coronary artery disease (as defined by the American College of Sports Medicine), that individual will not be permitted to take part in the study. Moreover, if a participant indicates he/she has greater than mild muscle or joint pain during the duration of the study, that individual will not be allowed to participate in the testing for that day. This test must then be rescheduled before the end of the four-week test period, or the data point will be treated as a missing data point.
A review of procedures specific to each test of aerobic power follows below:

Cooper’s 12-minute Run. This test will be administered on an indoor or outdoor track in the Corvallis or Greater Portland area. In the event that this test is administered outdoors, an effort will be made to administer the test under “good” (i.e. low wind, little precipitation) conditions. In the event that conditions are “poor,” the test will be rescheduled at a later time. In all cases, the track lane/lanes in which the test is administered will be free of obstacles, including other runners/walkers.

Participants will be instructed to complete as many laps on the track as possible during the 12-minute test period, with an emphasis on pacing oneself throughout the duration of the test. The test administrator will count the laps an individual completes during the 12-minute test period, while calling out the time elapsed at 3, 6, and 9 minutes and verbally encouraging the participants. At the end of the 12-minute period, the test administrator will call to the participants to stop, and will use a measuring wheel to determine the fraction of the last lap completed by each participant. This distance will then be added to the distance determined by the number of laps completed to give the total distance covered during the test. Cooper’s (1968) standardized equation will then be used to convert the distance run to an estimate of VO$_2$max.

Multistage 20-meter Shuttle Run (Leger protocol). The multistage 20-meter shuttle run will be administered on a gymnasium floor at Oregon State University or in the Greater Portland area. In this test, individuals will run a back-and-forth course (“shuttles”) between two cones placed 20 meters apart from one another. The speed of the run will start at 8.5 km/hr, and increase by 0.5 km/hr every minute. Participant
speed will be governed by a recorded, audible “beep” that will sound each time the participant reaches a cone to complete a shuttle. The number of shuttles per stage is coordinated with the speed of the beeps so that each stage is approximately a minute in length, and participants have the potential to complete up to 18 stages of successively faster shuttles. Participants will receive verbal encouragement throughout the duration of the test. When an individual fails to complete two successive shuttles in the allotted time between “beeps”, the test will be terminated and the individual’s estimated VO$_{2\text{max}}$ will be determined from a printed table by using the number of shuttles completed (Ramsbottom et al., 1988).

**Incremental treadmill test.** This testing will be completed in the Oregon State University Human Performance Laboratory, and will be conducted using a ParvoMedics TrueMax 2400 metabolic cart. During these tests, participants will first complete a 3-minute warm-up stage. The speed of the treadmill during this warm-up stage will be constant (5.0 to 7.0 mph, depending on the relative fitness of the individual) and the treadmill incline will be set at 0%. Following this warm-up period, the treadmill grade will be increased to 3.0% for the first of successive one-minute stages. At the completion of each stage, the treadmill speed will be increased by 0.5 mph. When the participant reaches an RPE of 13-14 (Borg scale), the test administrator will stop increasing treadmill speed and begin increasing the grade of the treadmill by 1% with each new stage. When the participant indicates that he/she cannot continue, the test will be terminated. A “true” maximal test will be completed if a participants reaches two of the following three criteria: (1) a VO$_2$ plateau evidenced by an oxygen uptake difference of less than 2.1 mL/kg from the previous
stage (Franklin, 2000), (2) an RER of greater than 1.15 (Franklin, 2000), or a maximal heart rate within ±10 beats of the age-predicted maximum (Poole et al., 2007). If a participant does not attain the requirements necessary to determine a maximal test, the participant will be given the option to re-test, as long as the repeated test falls within the four-week time period prescribed for each individual to complete the test protocol.

STATISTICAL ANALYSES

The following data analysis will be employed for specific research aims:

Research Aim 1. A two-random facet (instrument and occasion) G-study will be employed to answer the first research aim. In this two-facet crossed design, variance attributable to participant, test, and occasion will be observable, as well as shared variance between participant and test (p x t, indicating that participant characteristics may increase the variability of one test, but not the other), participant and occasion (p x o, indicating that participant characteristics may increase test variability on one or more occasions), and test and occasion (t x o, indicating that one test may be more variable than the other when repeated on more than one occasion). Finally, the magnitude of variance unexplained by the selected facets will also be able to be estimated.

A 2 x 3 (test by occasion) repeated measure ANOVA will be used to calculate the estimated variance components for each source of variability (t, o), as well as the interactions among variables (p x t, p x o, t x o). These estimated variance components will then be used to calculate the percentage of total variance in the sample data attributable to each facet and facet interaction pair.
Research Aim 2. A D-study will be performed to make recommendations as to how best reduce the variance in the measured aerobic power of an individual participating in a field test. Since the absolute value of VO_{2max} is important to coaches and physical educators, the D-study will use equations specific to “absolute decision” D-studies (Shavelson and Webb, 1991). This study will then examine the relative size of the error variance when using one or two field tests of aerobic power on one, two, or four different occasions. The relative magnitude of change in the error variance will then allow the researcher to recommend the ideal number of aerobic performance tests and test occasions necessary to achieve a reliable measure of aerobic power.

Research Aim 3. Test validity will be assessed using several methods:

1. VO_{2max} data from the individuals who completed the incremental treadmill test will be compared to the VO_{2max} data from the two field tests via ANOVA. If occasion-to-occasion (o) variability is relatively small, the data from the first administration of the field tests will be compared to the criterion score. This analysis using the initial scores is the preferred comparison, as it will remove any familiarization effect that will not be present with the single administration of the incremental treadmill test. However, if the occasion-to-occasion variability is large, the highest score from each field test will be compared to the criterion score. If this type of analysis is necessary, it will be assumed that familiarization with a test was particularly important for an optimal score, and that the highest score is, therefore, most indicative of that participant’s true score for that test.

In the event the ANOVA returns a signification F-value, a post-hoc Tukey’s analysis will be used to determine which of the field tests are significantly different
from the criterion incremental treadmill test. Pearson’s product-moment correlation coefficient will also be calculated between the incremental treadmill test data and the data from the two field tests to determine the magnitude of the relationship between the tests.

(2) Two Bland-Altman plots will be constructed comparing each field test with the incremental treadmill test. A 95% limit of agreement will be visually estimated from each plot.

(3) The magnitude of the instrument-to-instrument (t) variability determined from the G-study will be evaluated. If this source of variability is large, it is unlikely that both tests are valid in the measured population.
REFERENCES


