DEVELOPMENT AND CROSS-VALIDATION OF AEROBIC CAPACITY PREDICTION MODELS IN ADOLESCENT YOUTH

by

Ryan Donald Burns

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Exercise and Sport Science

The University of Utah

December 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Ryan Donald Burns**

has been approved by the following supervisory committee members:

James C. Hannon	, Chair	**8/26/14** Date Approved
Timothy A. Brusseau	, Member	**8/26/14** Date Approved
Barry Shultz	, Member	**8/26/14** Date Approved
Patricia Eisenman	, Member	**8/26/14** Date Approved
Matthew T. Mahar	, Member	**8/26/14** Date Approved

and by **James C. Hannon** , Chair/Dean of

the Department/College/School of **Exercise and Sport Science**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Cardiorespiratory endurance is a major component of health-related fitness testing in physical education. FITNESSGRAM recommends the 1-mile Run/Walk (1-MRW) or the Progressive Aerobic Cardiovascular Endurance Run (PACER) to assess cardiorespiratory endurance by estimating aerobic capacity, or $VO_{2\ Peak}$. No research to date has cross-validated prediction models from both 1-MRW and PACER using current FITNESSGRAM criterion-referenced (CR) standards. Additionally, new prediction models for 1-MRW without a body mass index (BMI) term are needed to attenuate the problems incorporating this index into an aerobic capacity model. The purpose of this dissertation was to cross-validate various prediction models using 1-MRW and PACER and to develop alternative 1-MRW aerobic capacity prediction models for adolescent youth. Participants included 90 students aged 13 to 16 years. Each student completed the 1-MRW and PACER, in addition to a maximal treadmill test to measure $VO_{2\ Peak}$. Multiple correlations among various models with measured $VO_{2\ Peak}$ were considered strong (R = 0.74 to 0.78). CR validity, examined using modified kappa (Kq), percentage of agreement (Pa), and *phi* was considered moderate among all models (Kq = 0.25 to 0.49; Pa = 72% to 79%; *phi* = 0.38 to 0.65). Two new models were developed from 1-MRW times, one linear and one quadratic model. The linear and quadratic models displayed multiple correlations of R = 0.77 and R = 0.82 with measured $VO_{2\ Peak}$, respectively. CR validity evidence was considered moderate with (Kq = 0.38; Pa = 73%; *phi* = 0.57) using the linear model and (*Kq* = 0.34; Pa = 70%; *phi* = 0.54) using the quadratic model. The accuracy of these models was confirmed using k-fold cross-validation. In conclusion, the prediction models demonstrated strong linear relationships with measured $VO_{2\ Peak}$, acceptable prediction error, and moderate CR agreement with measured $VO_{2\ Peak}$ using FITNESSGRAM's CR standards to categorize health groups. The new 1-MRW models displayed good predictive accuracy and moderate CR agreement with measured $VO_{2\ Peak}$ without using a BMI predictor. Despite evidence for predictive utility of the new models, they must be externally validated to ensure they can be generalizable to larger populations of students.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank several people who have helped me through my PhD program, supporting me personally and professionally throughout these 4 years. First, I want to thank my Mom and Dad who provided me the emotional support to go after my goals. Without them I would not have accomplished the difficult task of completing a doctorate, and their unconditional love has meant everything to me. Second, I want to thank my son Rylan who provided me the incentive and inspiration to persevere through these past 4 years. He has grounded me and constantly reminds me of what is important in life. Third, I would like to thank my committee: Drs. James Hannon, Timothy Brusseau, Patricia Eisenman, Barry Shultz, and Matthew Mahar. Each committee member has positively and uniquely contributed to the creation of this document and provided me valuable guidance in accordance to their expertise. I want to especially thank Dr. James Hannon. Dr. Hannon has been an excellent mentor these past 4 years. He has provided me numerous opportunities to grow professionally and has flawlessly guided me through the PhD program. Fourth, I would like to thank my fellow PhD students within the Department of Exercise and Sport Science. They have provided me help collecting data, provided guidance and entertainment during core Pedagogy classes, and most importantly provided me their friendship; I consider each my friend. Finally, I want to thank the Cooper Institute in Dallas, Texas, for funding this project that will hopefully advance the field of youth fitness testing.

CHAPTER 1

INTRODUCTION

Fitness assessment is a controversial aspect of physical education. The controversy arises from several factors that include deciding what fitness tests to implement for a respective fitness domain, how to administer tests within the time constraints of a physical education class, student reactions to the implemented fitness tests, and most importantly, how to interpret the scores so that a child can receive the maximum benefit from participating in the assessment (Cale, Harris, & Chen, 2007). The FITNESSGRAM fitness and physical activity assessment program is a significant advancement in youth fitness assessment and reporting. Instead of comparing a youth's score to a reference population to interpret fitness levels, as was the case employing the old Presidential Fitness program, FITNESSGRAM classifies students into one of three Healthy Fitness Zones by relating a fitness test score to a health-criterion measure (Welk, Going, Morrow, & Meredith, 2011). A child can use the Healthy Fitness Zone information to assess his or her own current health status and future health risk.

Despite the benefits of this program, potential exists for Healthy Fitness Zone misclassification (thus health-risk misclassification) that may lead to unnecessary negative emotional responses from youth, especially among adolescent youth who place body image and social acceptance of utmost importance to their well-being. Negative responses from youth may include decreases in self-esteem, self-efficacy, and a negative response for fitness testing (Cataldo, John, Chandran, Pati, & Shroyer, 2013). The following dissertation examined the validity of various field test prediction models for cardiorespiratory endurance in a sample of youth aged 13 to 16 years using current FITNESSGRAM standards. An attempt was also made to establish alternative aerobic capacity prediction models that are independent of a child's body mass index (BMI), which may yield equations with less inherent potential for misclassification in children who have relatively low or high BMI, but relatively low and high cardiorespiratory endurance, respectively. The partial reliance on BMI in classifying students may communicate incorrect messages to students in the context of cardiorespiratory endurance because of the limitations of this index. Accomplishing these objectives provided further insights into the potential for health-risk misclassification using FITNESSGRAM and provided new alternative models that may attenuate misclassification risk in specific groups of school-aged children.

History of Youth Fitness Assessment

Fitness testing is a common and important component of most physical education programs. Youth fitness testing in the US has an extensive history starting in the 1950s when Kraus and Hirschland's (1954) comparative study found that American youth were far less fit than their European counterparts. President Dwight D. Eisenhower, former Allied Commander during WWII, learned of this study and became increasingly concerned about the impact of poor fitness levels on the readiness of American youth for military service. As a result, the President's Council on Youth Fitness was established in

1956, which aimed to promote, encourage, and motivate American youth to become physically active and participate in sport. The link between physical fitness and military preparedness continued through the 1960s when fitness testing protocols focused on performance-related fitness, which was consistent with the growing sport culture in the US and "the more, the better" attitude regarding athletic performance (Morrow, Zhu, Franks, Meredith, & Spain, 2009). Although athletic performance is of interest to the student and teacher, many tests of athletic performance do not necessarily relate to health outcomes. Indeed, in many instances, students who were healthy did not meet expectations on these tests of athletic performance (Welk et al., 2011). However, starting in the 1970s, the better understanding of the relationship between fitness and health (Jackson, 2006) and the publication of *Aerobics* by Dr. Kenneth Cooper in 1968, among other factors, led to the emergence of the concept of *health-related physical fitness.*

Health-Related Physical Fitness

Health-related fitness is generally separated into five domains consisting of body composition, cardiorespiratory endurance, muscular strength, muscular endurance, and flexibility (National Association for Sport and Physical Education [NASPE], 2011). In the 1980s, obesity rates in the pediatric and adult populations were on the rise, and research yielded evidence of a link between excess adiposity, poor cardiorespiratory endurance, and increased chronic disease risk and mortality (Morrow et al., 2009). This research suggested that the domains of body composition and cardiorespiratory endurance seemed to have the strongest links to health; therefore, they were of great public interest.

Of the five domains of health-related physical fitness, cardiorespiratory endurance is considered by many one of the most important domains. Because of this, it is one of the most commonly assessed in physical education settings and is an important consideration when assessing health status in youth (Eisenmann, Welk, & Ihmels, 2007; Lee, Blair, & Jackson, 1999; Sui, Hooker, & Lee, 2007). Cardiorespiratory endurance, also known as aerobic fitness, cardiovascular fitness, or cardiorespiratory fitness, has strong links to cardiometabolic health. The construct of cardiorespiratory endurance is quantified using aerobic capacity, or $VO_{2\ Peak}$, which is the ability of the heart and lungs to circulate oxygenated blood to exercising tissues, the ability of the muscle cells to extract and use the oxygen for energy production, and the ability of the circulatory system to return blood back to the heart (NASPE, 2011). Recent evidence suggested that 42% of American youth aged 12 to 15 years have inadequate levels of cardiorespiratory endurance (Gahche et al., 2014). Welk, Laurson, and Eisenmann (2011) demonstrated that $VO_{2\ Peak}$ could be used to differentiate youth with and without metabolic syndrome, and other research supports that low levels of $VO_{2\ Peak}$ are associated with cardiovascular disease risk factors in adults (Blair, Goodyear, & Gibbons, 1984; Blair et al., 1989; Blair et al., 1995). Despite increased recognition of low cardiorespiratory endurance as a risk factor for adverse chronic medical conditions (Lobelo & Ruiz, 2007), little consensus on an acceptable classification system for the youth population exists.

Although much research supports that cardiorespiratory endurance is important to assess in children because of its links to laboratory health markers and mortality (Blair et al., 1989; Eisenmann et al., 2004; Freedman, Khan, Dietz, Srinivasan, & Berenson, 2001; Must, Jacques, Dallal, Bajema, & Dietz, 1992; Ogden, Carroll, Curtin, Lamb, & Flegal, 2010; Ortega et al., 2008), identifying students who may be at risk for developing chronic disease later in life is a challenge. The use of norm-referenced standards, which were often used in performance-related fitness assessment, cannot provide students with information regarding health outcomes because performance is assessed on how a student's fitness test score compares to a reference population. Therefore, along with the shift in paradigms from performance-related fitness to health-related fitness, a shift in how test scores are interpreted and reported to students has also occurred, from norm-referenced standards based on reference population percentile-ranks, to criterion-referenced standards based on how a score compares to an absolute health-related criterion (Zhu et al., 2010). The FITNESSGRAM program used in physical education settings uses the concept of criterion-referenced (CR) standards to identify and subsequently inform children who are at risk for developing health risk factors during childhood and adolescence.

Criterion-Referenced Standards

FITNESSGRAM is a fitness assessment and physical activity reporting program that provides physical education teachers with validated field-based fitness and physical activity assessments (Meredith & Welk, 2010). Appropriate uses of the fitness and activity assessments used in FITNESSGRAM include teaching self-monitoring skills, promoting educational outcomes, and most importantly to the students, providing personalized information about levels of health-related fitness (Welk, Going, Morrow, & Meredith, 2011). The FITNESSGRAM software generates printed reports of fitness test outcomes for each student that are in an easy-to-read format while additionally providing personalized suggestions based on test results. One of the other advantages to using the FITNESSGRAM program is the use of CR standards, which are based on how fit a child has to be in order to receive health benefits.

Using norm-referenced standards, which are employed in performance-based assessment, students are given a percentile rank on how they compare to a norm-referenced population. There are three major limitations to use of norm-referenced standards (Zhu, Mahar, Welk, Going, & Cureton, 2011). First, it is difficult to update norms regularly, as it takes a large and diverse sample to accurately develop an ethnically and physically representative population from which these fitness percentile ranks are based. Second, interpretation of a student's result is based on the referenced population; therefore, accurate generalization may be an issue in some instances. Third, norm-referenced standards often discourage children who are not "physically fit" and reward children who are.

One example of a norm-referenced standard in a youth fitness context was the use of the former Presidential Fitness Award Program. This program recognized students who reached the 85^{th} percentile on tests of physical fitness. The President's Council for Fitness, Sports, and Nutrition retired the President's Challenge fitness test in 2012 and adopted the FITNESSGRAM as the national youth fitness test in part because of the shift in emphasis on health outcomes, not athletic performance. Criterion-referenced standards differ from norm-referenced standards in that student performance is compared to an absolute criterion measure (a health outcome) as opposed to comparing a child's performance to his or her own peers. Therefore, information pertaining to current health and possible future health risk are reported to students based on their test scores within a respective health-related fitness domain.

FITNESSGRAM uses cut-off scores to classify students into three Healthy Fitness Zones: the Healthy Fitness Zone (HFZ), Needs Improvement-some risk (NI-some risk), and Needs Improvement-health risk (NI-health risk). These three fitness zones are an update from a previous dichotomous Healthy Fitness Zone categorization (HFZ, NI). A dichotomous classification method causes some confusion regarding health-risk interpretation, as there is not much difference in health outcomes among youth who have scores that lie just above or just below established cut-off scores (Welk et al., 2011). Therefore, the use of a three Healthy Fitness Zone classification method was deemed more appropriate. The HFZ indicates that a child has a level of fitness that is sufficient for good health. The NI-some risk zone indicates that a child has fitness scores that are close to NI-health risk, and that he or she should strive to improve their score to reach the HFZ for a specific fitness domain. Finally, the NIZ-health risk gives warning to children that their fitness levels may develop into potential health risk if they were to continue tracking at current levels.

Development of Fitness Zone Cut-Off Scores

Setting CR standards for youth fitness is a difficult task. This is because health risks are not as easily detected in youth as they are in adults, compounded with the fact that a significant amount of natural growth and maturation occurs during adolescence. The Healthy Fitness Zones used in the FITNESSGRAM program have cut-off scores that are age and gender specific; therefore, growth and maturation were taken into account when developing the current cut-off scores. However, developing standards requires a sequential process that is necessary to establish accurate Healthy Fitness Zone classification (Zhu et al., 2011) and necessary to understand when conducting research regarding the validity of the classification system.

The first step in developing CR standards is to select a health outcome measure or criterion measure from which the cut-off scores within each fitness domain will be based. Although no absolutely correct answer for selecting an outcome measure within a fitness context exists, FITNESSGRAM used the presence of the *metabolic syndrome* as the most appropriate outcome for the domains of body composition and cardiorespiratory endurance because metabolic syndrome is a major contributor in the development of chronic disease later in adulthood (Ford, 2005). Laboratory measures that indicate an individual has the metabolic syndrome include a) elevated waist circumference, b) elevated systolic and diastolic blood pressure, c) low fasting HDL-C, d) elevated fasting triglycerides, and e) elevated fasting blood glucose. If a youth met three or more of the five age and gender specific criteria for metabolic syndrome, he or she was classified as positive for having metabolic syndrome (Welk et al., 2011).

One of the main objectives of establishing criterion-referenced fitness standards is to determine the level of fitness within a specific domain that is needed for good health (Welk et al., 2011). When developing fitness standards, FITNESSGRAM uses the absence (testing negative for) of metabolic syndrome as “good” health. Therefore, when setting specific fitness domain standards a relationship must be established between the construct of interest (body composition, cardiorespiratory endurance) and metabolic syndrome status. The most accurate measurement of a construct is often referred to as a gold-standard measure; however, in physical education settings, acquiring gold standard measures for body composition (hydrostatic weighing, DEXA, MRI, CT scan, air displacement plethysmography) and cardiorespiratory endurance (directly measured $VO_{2\ Peak}$) is not feasible (Safrit, Baumgartner, Jackson, & Stamm, 1980). Therefore, reliable and validated field tests are used to estimate body composition and cardiorespiratory endurance.

FITNESSGRAM uses percent body fat estimated from two-site skinfold assessment (SKF) and body mass index (BMI) to evaluate body composition, and the tests of 1-mile Run/Walk (1-MRW), the Progressive Aerobic Cardiovascular Endurance Run (PACER), or the 1-mile Walk tests to estimate $VO_{2\ Peak}$ (Meredith & Welk, 2010). Age and gender specific cut-off scores for each of these two health-related fitness domains were developed using receiver operating characteristic curve (ROC curve) analyses (Zhu et al., 2011).

When classifying students on whether or not they are at risk for developing metabolic syndrome (the criterion health measure), there are four possible outcomes. A *true positive* is when both the criterion measure and a field test indicate that a student has achieved a standard that he/she may have metabolic syndrome. A *true negative* is when both the criterion measure and a field test indicate a student has not met the standard for having metabolic syndrome. A *false positive* is when a field test incorrectly identifies a student as having met the standard of having metabolic syndrome when he or she does not, determined by the criterion. Finally, a *false negative* is when a field test incorrectly identifies a student having not met the standards of having metabolic syndrome when he or she actually does, as determined by the criterion. ROC curve analysis uses these four possible outcomes for a given field test to calculate the *sensitivity* and 1-*specificity* of that field test (Pagano & Gauvreau, 2000).

Sensitivity is the probability that a student tests positive based on a field assessment (T^+) given that he or she actually has a condition (D), or $P(T^+| D)$. *Sensitivity* is synonymous with the probability of achieving a *true positive* for a given field test. *Specificity* is the probability that a student tests negative (T^-) based on field test performance given that he/she does not have metabolic syndrome, $P(T^-|D^-)$, or a *true negative* (Pagano & Gavreau, 2000); (1-specificity) therefore is the probability of achieving a false positive based on field test performance given that a student does not have metabolic syndrome, or $P(T^+|D^-)$, a *false positive*.

Using ROC curve analysis, *sensitivity* is plotted against (1-*specificity*) or the probability of achieving a *true positive* against the probability of achieving a *false positive* for a given field test across a range of possible cut-off scores. To establish the most effective cut-off score for a given field test, *sensitivity* and *specificity* would have to be maximized; this would maximize true positive rates while minimizing false-positive rates for a field test (see Figure 1.1). A standard or cut-off score that maximizes *sensitivity* and *specificity* is the cut-off score that would be the most effective for classifying students into those who do and those who do not have a certain disease or condition (metabolic syndrome) as determined by the criterion measure (see Figure 1.1; Pagano, 2000).

FITNESSGRAM used the aforementioned ROC curve techniques to develop body composition and cardiorespiratory endurance cut-off scores for a three Healthy Fitness Zones classification scheme. However, in order to establish three Healthy Fitness

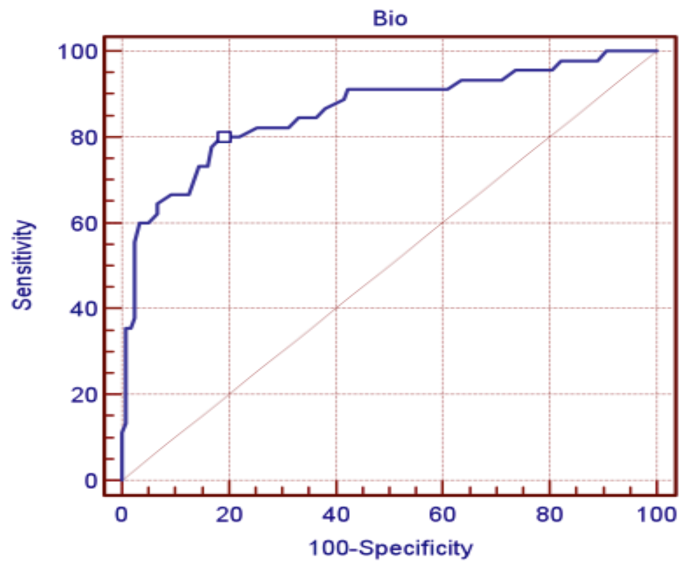

Figure 1.1. A sample ROC curve plot depicting sensitivity on the y-axis and (1-specificity) on the x-axis. Optimal threshold for dichotomous classification is indicated by the square datum on the upper left section of the curve (maximizing sensitivity and specificity).

Zones for a respective fitness domain, two thresholds had to be identified. Using cardiorespiratory endurance as an example, one cut-off score emphasized *sensitivity* over *specificity*, the low risk or HFZ cut-off, ensuring that children with metabolic syndrome would have aerobic capacity scores below this threshold, and one cut-off score emphasized *specificity* over *sensitivity*, the high risk or NI cut-off, ensuring that children without metabolic syndrome would have aerobic capacity scores above this threshold. This same analytic technique was also used for the establishment of body composition standards (Laurson, Eisenmann, & Welk, 2011). To partially account for natural growth and maturation in the analyses, fitness scores were converted to *z*-scores using a LMS statistical technique with skewness (L), median (M), and coefficient of variation parameters (S; Welk et al., 2011). These *z*-scores were then used in the ROC curve analyses to develop the specific fitness test thresholds.

Field Test Equating

Ideally, one standardized field test would be used for a fitness domain (i.e., PACER for cardiorespiratory endurance, SKF for body composition, etc.), as there may be inconsistency in scoring if an assessment program used a multiple test approach. However, the single test approach is not very feasible in reality because of practical and historical issues; one school may be used to conducting a certain test for cardiorespiratory endurance (e.g., 1-MRW) or there may be a lack of space or time constraints to conduct certain other field test (e.g., SKF assessment). Therefore, a multiple field test approach is used by FITNESSGRAM to accommodate these issues in physical education settings (Safrit, Baumgartner, Jackson, & Stamm, 1990). Test equating is used to put two or more tests that measure the same construct onto the same scale for comparison. A prime example is equating the PACER test score (in laps) to 1-MRW times (in minutes). Both of these field tests measure cardiorespiratory endurance, but they are measured on different scales. Zhu, Plowman, and Park (2010) developed the Primary Field Test Centered Equating Method to equate PACER scores with 1-MRW times (see Figure 1.2). During this procedure, one field test is used as a primary field test and one as a secondary field test. The 1-MRW times are used as the primary field test because of its established moderate to strong associations with $VO_{2\ Peak}$, the criterion or gold-standard measure for cardiorespiratory endurance. Figure 1.2 from Zhu et al. (2010) visually depicts the systematic approach used in field test equating. Once field tests are equated, then the secondary field test such as the PACER test can be converted to 1-MRW times (referred to as Mile PEQ scores), which can then be used to estimate aerobic capacity. FITNESSGRAM uses the Cureton, Sloniger, O'Bannon, Black, and McCormack (1995)

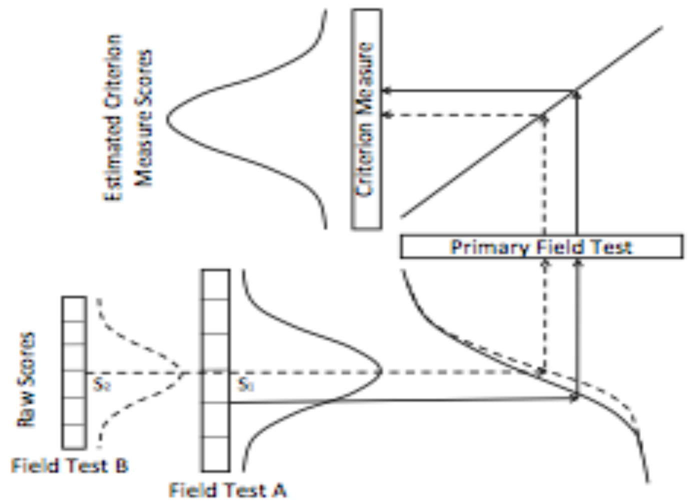

Figure 1.2. Conceptual illustration of the primary field-test centered equating method for setting cut-off scores (from Zhu, Plowman, & Park, 2010).

model to predict $VO_{2\ Peak}$ from 1-MRW or Mile PEQ times:

$$VO_{2\ Peak} = 0.21(\text{Age} \times \text{Gender}) - 0.84(\text{BMI}) - 8.41(\text{Time}) + 0.34\ (\text{Time})^2 + 108.94 \qquad [1]$$

where 0 = girl and 1 = boy for Gender,

Age = chronological age in years, and

Time = minutes.

Until recently, the aforementioned equating method has been used by FITNESSGRAM to predict $VO_{2\ Peak}$ from PACER performance. However, Mahar et al. (2014) developed a new model that more directly estimates $VO_{2\ Peak}$ using only PACER laps and age as predictor variables. The study validating this model is currently in preparation for publication.

Statement of the Problem

The new CR standards developed by FITNESSGRAM for the health-related fitness domain of cardiorespiratory endurance were developed using a large National Health and Nutrition Examination Survey (NHANES) database from measures collected between the years 1999 to 2004 that included an ethnically diverse sample of approximately 1,240 children aged 12 to 18 years (Welk et al., 2011). Standards were developed relating the health criterion "metabolic syndrome" with a submaximal aerobic capacity treadmill test to estimate $VO_{2\ Peak}$ based on heart rate response. The relationship between estimated $VO_{2\ Peak}$ and the health outcome was used to establish cut-off scores for classification into HFZ, NI-some risk, and NI-health risk fitness zones.

The validity of various 1-MRW and PACER prediction models to estimate measured $VO_{2\ Peak}$ using an independent sample has been limited in adolescent school-aged children aged 13 to 16 years. Additionally, limited research has examined the CR validity, or the consistency in classification between measured and estimated $VO_{2\ Peak}$ into Healthy Fitness Zones, among various 1-MRW and PACER $VO_{2\ Peak}$ prediction models using current FITNESSGRAM CR standards.

Finally, the prediction model developed by Cureton et al. (1995), which estimates aerobic capacity from 1-MRW, introduces an issue in classifying youth into appropriate Healthy Fitness Zones. Although not without inherent limitations, BMI is often used to assess body composition when using the FITNESSGRAM program because of its ease of administration and calculation. The Cureton et al. model used to estimate $VO_{2\ Peak}$ contains a negative BMI coefficient (see equation 1). Therefore, a child who has a higher BMI will essentially be penalized twice when using FITNESSGRAM assessment, once for the body composition assessment (when using BMI) and once for cardiorespiratory endurance by lowering his or her estimated $VO_{2\ Peak}$ score due to the presence of the negative BMI term. Given the limitations of BMI, this raises a potential significant issue when attempting to classify children who have low or high BMI by either overestimating or underestimating relative aerobic capacity because of the inability of BMI to distinguish between fat mass (FM) and fat free mass (FFM). Additionally, the collection of height and weight information used to calculate BMI may present problems for some physical education teachers if these measures are not readily available. Therefore, to circumvent these issues, an attempt should be made to develop a new aerobic capacity prediction model from 1-MRW times without a BMI predictor. It must be determined if BMI is needed to yield a model with good predictive accuracy and CR validity used to estimate $VO_{2\ Peak}$ in adolescents aged 13 to 16 years.

Research Questions

The following research questions were addressed in this dissertation:

1. What is the evidence for the external norm-referenced validity and CR validity of various 1-MRW and PACER aerobic capacity prediction models to estimate measured $VO_{2\ Peak}$ using current FITNESSGRAM cut-off scores in adolescents?

2. Can an aerobic capacity prediction model without a BMI term have good predictive accuracy in estimating an adolescent's measured $VO_{2\ Peak}$ from 1-MRW times and improve CR classification agreement with measured $VO_{2\ Peak}$ into the Healthy Fitness Zones?

Study Purpose

The purposes of this project were to (a) examine the external norm-referenced validity and CR validity of various aerobic capacity prediction models to estimate measured $VO_{2\ Peak}$ in adolescent youth using current FITNESSGRAM cut-off scores and (b) develop aerobic capacity prediction models without a BMI predictor from 1-MRW times in adolescents.

It was hypothesized that there would be moderate-to-strong multiple correlations (R = .60 to .80) between various $VO_{2\ Peak}$ prediction models with measured $VO_{2\ Peak}$ and moderate classification agreement (kappa = .40 to .60; proportion of agreement = 60% to 80%) between measured and estimated $VO_{2\ Peak}$ into a three Healthy Fitness Zones scheme. It was also hypothesized that the development of a new 1-MRW aerobic capacity prediction model would result in a similar coefficient of determination (R^2) and standard error of estimate (SEE) compared to the current Cureton et al. (1995) model, suggesting similar predictive accuracy of measured $VO_{2\ peak}$. The strong predictive accuracy of the developed models without BMI will also yield acceptable CR agreement with measured $VO_{2\ Peak}$ into Healthy Fitness Zones. Therefore, the developed models will give evidence for strong utility in school settings due to the elimination of the practical issues that manifest using a BMI predictor.

Significance

FITNESSGRAM is a widely used fitness and physical activity assessment program in physical education curricula. The Presidential Youth Fitness Program has adopted FITNESSGRAM's assessment protocols, which includes the use of the current

CR standards for student classification. Therefore, research examining the validity of these standards has the potential to impact millions of youth relative to how they are classified based on health-related fitness test performance.

The new FITNESSGRAM CR standards were developed to classify youth into one of three Healthy Fitness Zones: HFZ, NI-some risk, and NI-health risk. The CR validity of prediction models classifying youth into these Healthy Fitness Zones needs to be examined by assessing agreement with a laboratory measure. A high level of agreement (agreement ≥ 80%) between a prediction model's estimate and the laboratory measure will help further establish the model's accuracy for identifying youth who are at high or low risk for developing chronic disease. However, low levels of agreement (agreement ≤ 60%) between a prediction model's estimate and the laboratory measure will suggest that the cut-off scores or the prediction equations that estimate the construct may need to be revised in order for FITNESSGRAM to be a more valid classification program. Additionally, the development of an alternative prediction model to estimate aerobic capacity may yield an equation with greater predictive utility and less inherent potential for misclassification compared to the current Cureton et al. (1995) model. Using predictor variables that do not include a BMI coefficient may attenuate possible misclassification for youth that have a low or high BMI, but relatively low and high cardiorespiratory endurance, respectively.

Assumptions

The following assumptions were made prior to conducting this project:

1. It was assumed that students gave maximal effort during field test assessments

(1-MRW and PACER) and the laboratory assessment (measured $VO_{2\ Peak}$).

2. It was assumed that the field test environment (different schools, different running surfaces, etc.) did not significantly affect performance.

3. It was assumed that there were no significant cardiorespiratory endurance changes within the 3-week data collection window.

Delimitations

The following delimitations were made prior to conducting this project:

1. Subjects were delimited to adolescent youth aged 13 to 16 years.

2. Subjects were recruited from private schools located in the Salt Lake Valley.

3. For a given youth, all cardiorespiratory endurance data were collected within 3 weeks.

4. At least a 48-hour recovery was given between field tests and the laboratory test.

5. The PACER test was completed indoors for all students.

6. The 1-MRW was completed outdoors for all students.

7. For each student, measured $VO_{2\ Peak}$ was obtained after both field tests were completed.

Limitations

This project had the following limitations that must be considered before interpreting results:

1. Results of this study cannot be generalized to younger or older age groups.

2. There were an unequal number of boys and girls in the sample.

3. Slight weather changes could have affected the outcome of the 1-MRW.

4. Youth performed cardiorespiratory field tests in different group sizes; this was done for convenience to reduce subject burden (scheduling conflicts).

Definition of Terms

Aerobic capacity is operationally defined as $VO_{2\ Peak}$ or the maximal amount of oxygen the body can take-in, transport, and utilize during maximal, dynamic, and large muscle exercise − typically expressed as ($mL \cdot kg^{-1} \cdot min^{-1}$).

Cardiorespiratory endurance is a physical characteristic that is a key component of health-related fitness assessment in school-aged children due its links to cardio-metabolic health (also referred to as cardiorespiratory fitness, aerobic fitness, or cardiovascular fitness).

Criterion-referenced (CR) is the evaluation of a test-takers' performance by comparing it to an absolute "criterion" standard (health outcome) as opposed to a "norm-referenced" standard.

Criterion-referenced (CR) validity is the extent to which measured and estimated measures agree in classifying individuals into established groups (i.e., Healthy Fitness Zones) developed using criterion-referenced cut-off scores.

External validity is the extent to which results from one study are supported in another study using a different sample.

FITNESSGRAM is a youth fitness and physical activity assessment program used in school-based physical education programming to facilitate the collection and processing of youth fitness and physical activity data.

Healthy Fitness Zone (HFZ) is a fitness zone indicating that a child has achieved a sufficient level of fitness for good health.

Needs Improvement Zone (NIZ) is a fitness zone indicating that a child has a lower level of fitness that may lead to potential health risks if he or she were to continue to track at that level. NIZ is divided into NIZ-some risk and NIZ health-risk subzones.

One-mile run/walk test (1-MRW) is the alternative field test for assessing cardiorespiratory endurance using the FITNESSGRAM battery.

Progressive aerobic cardiovascular endurance run (PACER) is a multistage cardiorespiratory endurance field test that is the recommended (default) assessment using FITNESSGRAM.

CHAPTER 2

CROSS-VALIDATION OF AEROBIC CAPACITY PREDICTION MODELS IN ADOLESCENTS

Introduction

Due to the declining health status of the US population, valid health-related fitness assessment is needed to identify youth at risk for developing chronic disease so that measures can be taken to prevent increased morbidity and mortality. The FITNESSGRAM program provides teachers and students with important information regarding indicators of health-related fitness that are associated with health outcomes in the pediatric population. FITNESSGRAM uses criterion-referenced (CR) standards that are based on how physically fit one would need to be in order to achieve health benefits (Welk et al., 2011). A health-related fitness domain that associates strongly with health outcomes is cardiorespiratory endurance (Blair et al., 1989; Eisenmann et al., 2004; Freedman et al., 2001; Must et al., 1992; Ogden et al., 2010; Ortega et al., 2008).

Cardiorespiratory endurance, also called cardiovascular fitness, aerobic fitness, or cardiorespiratory fitness, is considered to be one of the most important domains of health-related fitness because of its links to cardiometabolic health in both the pediatric and adult populations (Boreham & Riddoch, 2001; Eisenmann, 2004). Cardiorespiratory endurance can be operationally defined by aerobic capacity or $VO_{2\ Peak}$, which is the maximal amount of oxygen one can take-in, transport, and utilize during exercise. Research has shown that as many as one-third of American children have inadequate levels of cardiorespiratory endurance (Carnethon, Gulati, & Greenland, 2005; Pate et al., 2006) and that cardiorespiratory endurance tracks reasonably well from childhood through adolescence (Malina, 1996). Therefore, children who have low cardiorespiratory endurance tend to have low levels throughout their adolescent years and into adulthood.

FITNESSGRAM uses the 1-mile run/walk test (1-MRW) or the Progressive Aerobic Cardiovascular Endurance Run (PACER) to assess cardiorespiratory endurance. The 1-MRW is a widely-used field test to predict aerobic capacity with previous research demonstrating a moderate-to-strong relationship with measured $VO_{2\ Peak}$, having correlation coefficients ranging from r = -.56 to -.80 and test-retest intraclass reliability coefficients of R > .90 (Buono, Roby, Miale, Sallis, & Shepard, 1991; Burke, 1976; Cureton, Boileau, Lohman, & Misner, 1977; Kearney & Bynes, 1974). Estimated $VO_{2\ Peak}$ is calculated from 1-MRW times using the Cureton et al. (1995) model (Equation 1).

Although the 1-MRW is still widely used as a cardiorespiratory endurance field test, it is the PACER test, developed by Leger and Lambert (1982), which is the recommended (default) cardiorespiratory endurance test used in FITNESSGRAM. For a number of reasons, the PACER is recommended over 1-MRW in children. First, students are more likely to have a positive experience performing the PACER due to synchronous group running, the progressive nature of the test, and the use of background music. Second, the PACER helps students learn the skill of pacing. And third, students who have a poorer performance on the PACER will finish first and not be subjected to the embarrassment of being the last person to complete the test (Meredith & Welk, 2010).

The PACER test has correlation coefficients ranging from r = .60 to .87 with measured $VO_{2\ Peak}$ (Leger, Mercier, Gadourey, & Lambert, 1988; Mahoney, 1997) and has demonstrated test-retest reliability in school-aged children with intraclass coefficients ranging from R = .82 to .93 (Liu, Plowman, & Looney, 1992; Vincent, Barker, Clarke, & Harrison, 1999).

When estimating $VO_{2\ Peak}$ from PACER performance, PACER laps can be converted to 1-MRW times by use of the Primary Field Test Centered Equating Method (Zhu et al., 2010). The converted 1-MRW scores from PACER performance, referred to as Mile-PACER equated scores (Mile PEQ), are then used to estimate $VO_{2\ Peak}$ by use of the Cureton et al. (1995) model. Therefore, there is a double conversion to obtain a predicted $VO_{2\ Peak}$ score from PACER performance. However, other studies have produced additional models to estimate $VO_{2\ Peak}$ more directly from PACER performance (Mahar, Guerieri, Hanna, & Kemble, 2011).

The following Equations (2 and 3) are models developed from Mahar et al. (2011) that can be effectively used to estimate $VO_{2\ Peak}$ from PACER laps. A new PACER model has been developed (Equation 4), which was recently adopted by FITNESSGRAM as the model to estimate $VO_{2\ Peak}$ from PACER performance (Mahar et al., 2014). Equation 2 is a linear prediction model (Mahar Linear 2011), Equation 3 is a quadratic prediction model (Mahar Quadratic 2011), and Equation 4 is the current model employed by FITNESSGRAM to estimate $VO_{2\ Peak}$ from PACER performance (New PACER 2014):

$$VO_{2\ Peak} = 0.21(\text{Laps}) + 4.27(\text{Gender}) + 0.79(\text{Age}) - 0.79(\text{BMI}) + 40.35 \qquad [2]$$

$$VO_{2\ Peak} = 0.49(\text{Laps}) - 0.0029(\text{Laps})^2 - 0.62(\text{BMI}) + 0.35(\text{Gender} \times \text{Age}) + 41.77 \qquad [3]$$

$$VO_{2\ Peak} = 0.353(Laps) - 1.121(Age) + 45.619 \quad [4]$$

where 0 = girl and 1 = boy for Gender,

Age = chronological age in years,

BMI = Body Mass Index in $kg \cdot m^{-2}$, and Laps = number of Laps.

Following the conversion of cardiorespiratory field test scores to estimated $VO_{2\ Peak}$, students are then classified into one of three Healthy Fitness Zones. Using a three Healthy Fitness Zone classification scheme, messages to students can be delivered with more accuracy compared to a two Healthy Fitness Zone scheme because practically there are negligible differences in health risk for students who lie just above or below a single cut-off score. The three Healthy Fitness Zones include the Healthy Fitness Zone (HFZ), Needs Improvement-some risk (NI-some risk) subzone, and Needs Improvement-health risk (NI-health risk) subzone. Although field tests used to estimate $VO_{2\ Peak}$ were developed and cross-validated against measured $VO_{2\ Peak}$, there is limited research using an independent sample of adolescent youth aged 13 to 16 years comparing the predictive accuracy of various models from both 1-MRW and PACER scores with laboratory measured $VO_{2\ Peak}$ using current FITNESSGRAM standards.

A criterion test is the most valid measure of an assessed construct, in this case cardiorespiratory endurance. The gold-standard criterion measure for cardiorespiratory endurance is a graded maximal exercise $VO_{2\ Peak}$ test in a laboratory setting. Strong multiple correlations and low prediction error between a prediction model's estimate and criterion measure establishes strong evidence for the validity of that model as an estimate of the construct of interest. Weak multiple correlations and high prediction error suggests the validity of the field test itself is unacceptable in estimating a construct or may suggest that the prediction model used to estimate the construct may need revision.

When examining criterion-referenced (CR) validity evidence, estimated and measured $VO_{2\ Peak}$ scores are used to compare the resultant health-risk classifications into Healthy Fitness Zones. Despite the latest research showing improved CR agreement between the two field tests of cardiorespiratory endurance (Welk et al., 2011), no research to date has examined CR validity of both the 1-MRW and PACER models with measured $VO_{2\ Peak}$ using current standards in adolescent youth aged 13 to 16 years. Therefore, the purpose of this study was to examine the external validity of various 1-MRW and PACER $VO_{2\ Peak}$ prediction models against measured $VO_{2\ Peak}$ in a sample of adolescent youth aged 13−16 years. CR validity was also examined by comparing CR agreement of the placement into Healthy Fitness Zones using FITNESSGRAM's current CR standards. It was hypothesized that there will be moderate-to-strong multiple correlations (R = .60 to .80) and moderate CR agreement (percentage of agreement = 60% to 80%; kappa statistic = .40 to .60) between the various prediction models' estimates of $VO_{2\ Peak}$ and measured $VO_{2\ Peak}$ in this sample of adolescent youth aged 13 to 16 years.

Methods

Participants and Setting

Participants were 90 adolescent youth (38 girls, 52 boys) recruited from middle and high schools located in a metropolitan area in the Mountain West region of the United States. Data collection took place in both laboratory and field settings. The laboratory setting included having the students report to the University Human Performance Research Laboratory (HPRL) after school hours or on weekends for VO_2

$_{Peak}$ testing. Field settings included testing students on school grounds or at the University gymnasium or outdoor track. Written assent was obtained from the students and written consent was obtained from the parents prior to data collection. The University's Institutional Review Board (IRB) approved the protocols used in this study.

Tests of Cardiorespiratory Endurance

Laboratory measure: Maximal graded-exercise treadmill $VO_{2\ Peak}$ test. $VO_{2\ Peak}$ was measured using a maximal graded treadmill test to exhaustion. Previous research has suggested that determination of youth $VO_{2\ Peak}$ is as reliable as determining $VO_{2\ Peak}$ in adults with error across three tests within ± 4% (Welsman, Bywater, Farr, Welford, & Armstrong, 2005). Prior to each $VO_{2\ Peak}$ test, gas analyzers were calibrated for expired air with certified gases of known standard concentrations (4.00% CO_2, 16.00% O_2). Volume calibration was employed using a 3-Liter calibration syringe (Hans Rudolph, Kansas City, MO, USA). Gas and flowmeter calibration was repeated until error was less than 3%. During the test, VO_2 and VCO_2 were measured continuously via open circuit spirometry and analyzed with the use of the ParvoMedics 2400 metabolic measurement system (Sandy, UT, USA).

Upon reporting to the HPRL, students were familiarized with the exercise equipment and then practiced running on the treadmill for approximately 2 minutes without holding onto the railing. The students then followed a progressive, maximal treadmill protocol appropriate for the youth population described by Mahar et al. (2011). The protocol proceeded in gender-specific, progressive, and incremental work stages. For girls, the treadmill speed increased to 5.0 mph within the first minute at 0% grade and this speed was maintained for the duration of the test. For boys, the treadmill speed increased to 5.5 mph within the first minute at 0% grade and this speed was maintained for the duration of the test. For both girls and boys, at the beginning of the second minute, the treadmill grade was raised to 2% and every minute thereafter, the grade increased by another 2% until the students were no longer able to continue. The students were able to voluntarily stop at any time by tapping the treadmill railing to communicate to the test administrator to stop the test. Meeting two of three criteria determined if a successful $VO_{2\ Peak}$ test was performed: (a) showing signs of intense effort (heavy breathing, facial flushing, unsteady gait, and sweating), (b) a heart rate ≥ 90% age-predicted maximum, and (c) a respiratory exchange ratio (RER) ≥ 1.0. $VO_{2\ Peak}$ was recorded in absolute ($L \cdot min^{-1}$) and relative ($ml \cdot kg^{-1} \cdot min^{-1}$) terms.

Field measure: Progressive aerobic cardiovascular endurance run (PACER). The PACER is the recommended cardiorespiratory endurance test used in the FITNESSGRAM assessment program. The PACER was administered indoors for all students on a marked gymnasium floor with background music and cadence given by an audio CD. Students were instructed to run from one floor marker to another marker set 20-m apart while keeping pace with a prerecorded cadence. A single beep sounded at the end of the time allotted for each lap. A triple beep sounded when the students had completed a stage of the test and indicated that the pacing would get progressively faster. The test was terminated when a student twice failed to reach the opposite marker in the allotted time frame or when he/she voluntarily stopped. Final score was recorded in completed "Laps" (Meredith & Welk, 2010).

Field measure: 1-mile run/walk test (1-MRW). The 1-MRW is one of the alternative cardiorespiratory endurance tests used in the FITNESSGRAM testing battery. The 1-MRW was administered on either a standard rubber track on school grounds or at the University's track facilities after school hours or on the weekends. All 1-MRW tests were conducted outdoors. Students were instructed to run and/or walk 1 mile as fast as possible. Time was kept via a handheld stopwatch (Robic Oslo M427; Oxford, CT, USA) and scored in minutes (Meredith & Welk, 2010).

Study Procedures

Data collection occurred on three separate testing days in both laboratory and field settings. A simple procedural flow chart is presented in Figure 2.1. Day 1 consisted of students completing the PACER test on school grounds during their physical education class or at the University's facilities. Day 2 occurred 1 week following the PACER test and consisted of having students complete the 1-MRW on school grounds during their physical education class or at the University's facilities. Day 1 and Day 2 testing was counterbalanced to control for a confounding order effect. Field tests were conducted in the afternoon during the last two class periods of the day or on the weekends at the University facilities. On Day 3, students reported to the HPRL either after school hours or on weekends to complete the $VO_{2\ Peak}$ test. $VO_{2\ Peak}$ testing occurred no less than 48 hours after the 1-MRW or PACER to allow for full recovery and no more than 2 two weeks after the 1-MRW or PACER to minimize changes in cardiorespiratory endurance levels. Prior to the treadmill test, each student's height and weight were recorded using a stadiometer (Seca 213; Chino, CS, USA) and medical scale (Tanita HD-314; Arlington Heights, IL, USA). Percent body fat (%BF) was also obtained using the Slaughter et al.

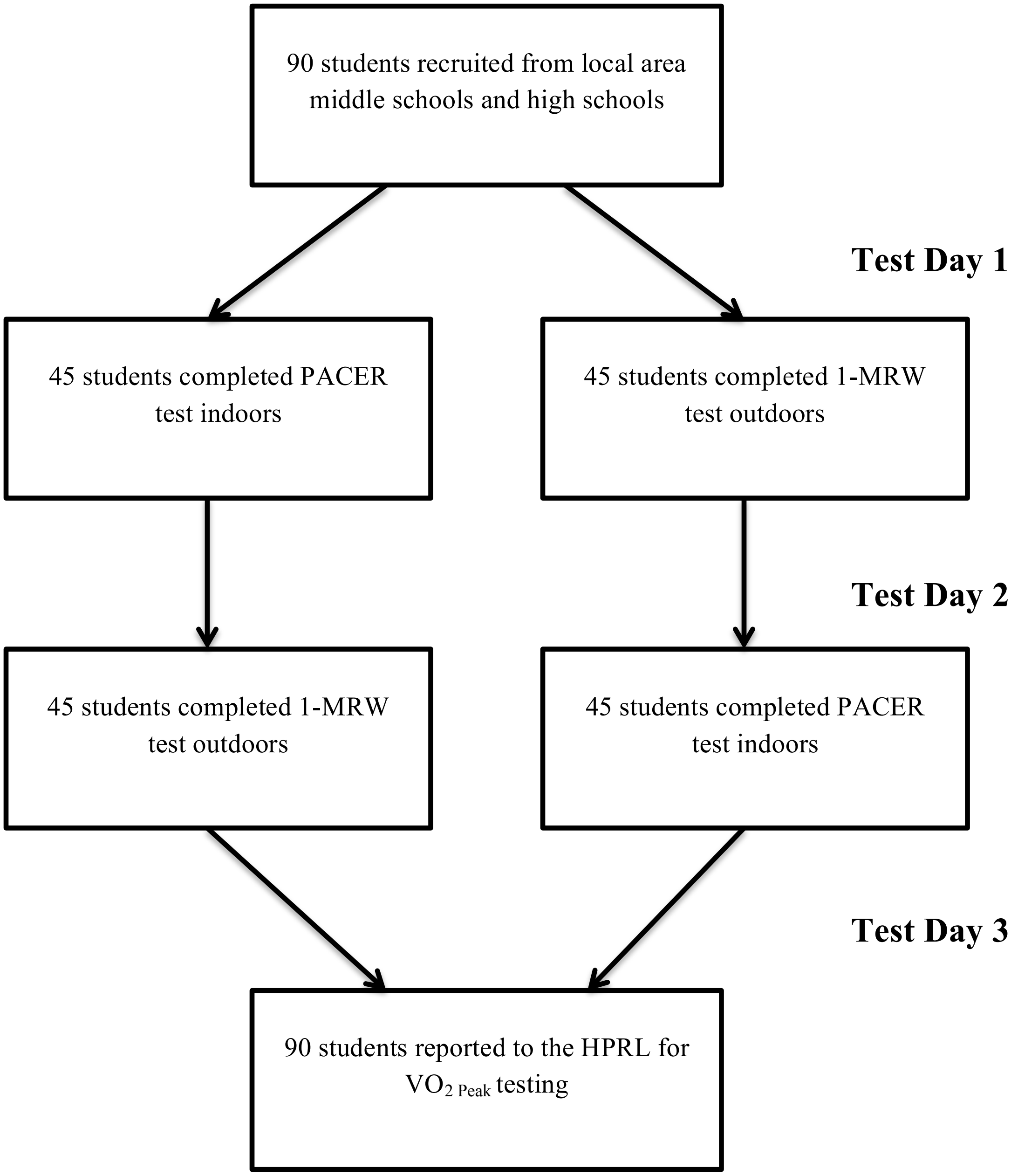

Figure 2.1. Procedure flow chart for collection of cardiorespiratory endurance data.

(1988) equation via two-site skinfold thickness assessment using a Lange skinfold caliper (Ann Arbor, MI, USA). Students were instructed to wear comfortable gym attire and to not have eaten within 2 hours prior to testing. In order to maintain testing consistency, all data were collected by the same trained graduate student in both laboratory and field settings.

Data Analysis

Data were screened for outliers and normality checked prior to any of the main analyses using k-density plots and the Shapiro-Wilk test for normality. Descriptive analyses were performed comparing gender and age differences in the anthropometric and fitness data using multiple 2 X 4 factorial ANOVA tests. A Bonferroni post hoc analysis was employed if a significant age main effect was found. Alpha level was adjusted appropriately for post hoc analyses using the Bonferroni adjustment.

The 1-MRW and PACER scores were converted to estimated $VO_{2\ Peak}$ using equation 1 for the 1-MRW scores, equations 2, 3, and 4 for the PACER scores, and also the Mile-PEQ method for PACER scores. Therefore, each student had five estimated $VO_{2\ Peak}$ scores, one from 1-MRW times and four from PACER laps using the aforementioned models. Validity was analyzed using multiple correlations, paired *t*-tests, and the total error (root mean square error) given by the following equation:

$$\text{Total Error} = \sqrt{}\ \Sigma(Y - Y') / N \qquad [5]$$

where Y is measured $VO_{2\ Peak}$,

Y' is estimated $VO_{2\ Peak}$,

and N is the total sample.

Individual agreement between measured and estimated $VO_{2\ Peak}$ for each model was examined using Residual vs. Fitted plots with estimated $VO_{2\ Peak}$ plotted on the x-axis and the residuals (Measured $VO_{2\ Peak}$ – Estimated $VO_{2\ Peak}$) plotted on the y-axis (O' Connor, Mahar, Laughlin, & Jackson, 2011). The mean differences (MD) and correlations between the $VO_{2\ Peak}$ residuals and estimates were reported for each examined model. CR agreement into the three Healthy Fitness Zone scheme was analyzed by comparing the classification from estimated $VO_{2\ Peak}$ with the classification from measured $VO_{2\ Peak}$ using percentage agreement (Pa), weighted kappa statistics (Kq), and a phi coefficient (ϕ). Proportion of Agreement into Healthy Fitness Zones was considered poor if below 60%, moderate if 60% to 80%, and excellent if above 80% (Hartmann, 1977). Statistical analyses were conducted using STATA v12.0 (College Station, TX, USA) statistical software.

Results

The descriptive, anthropometric, and fitness performance data for the total sample and within each gender group are presented in Table 2.1. Boys in this sample were taller than girls ($P < .05$), had lower %BF compared to girls ($P < .05$), and displayed higher absolute and relative measured $VO_{2\ Peak}$, faster 1-MRW times, and greater PACER laps compared to girls ($P < .05$). There were no statistically significant differences between genders in age (years), weight (kg), or BMI ($kg \cdot m^{-2}$). Statistically significant age main effects were present for height (m; $F_{(3,85)} = 23.17$, $P < .001$), weight (kg; $F_{(3, 85)} = 15.32$,

Table 2.1. Descriptive data for the total sample and within sex groups.

	Total (*N* = 90)		Girls (*n* = 38)		Boys (*n* = 52)	
	Mean	SD	Mean	SD	Mean	SD
Age (years)	14.68	1.27	14.61	1.22	14.72	1.32
Height (m)	1.69	0.12	1.63	0.05	**1.73**†	**0.14**
Weight (kg)	59.65	15.58	55.55	12.27	**62.79**†	**17.15**
BMI ($kg\cdot m^{-2}$)[1]	20.59	3.44	20.57	3.42	20.60	3.50
% Body Fat	20.80	7.68	**25.80**†	**6.97**	16.97	5.78
1-MRW (min)[2]	8.12	1.98	**9.29**†	**2.01**	7.22	1.42
PACER (Laps)[3]	57.52	26.20	41.48	18.37	**69.78**†	**24.73**
$VO_{2\,Peak}$ ($L\cdot min^{-1}$)	2.69	0.87	2.22	0.51	**3.05**†	**0.92**
$VO_{2\,Peak}$ ($ml\cdot kg^{-1}\cdot min^{-1}$)	45.23	8.16	39.95	4.80	**49.26**†	**7.91**

[1] BMI stands for Body Mass Index
[2] 1-MRW stands for the 1-mile Run/Walk test
[3] PACER stands for the Progressive Aerobic Cardiovascular Endurance Run
† Denotes statistically significant differences between genders, $P < .001$

$P < .001$), BMI ($F_{(3,85)} = 5.39$, $P < .001$) , and absolute measured $VO_{2\ Peak}$ ($F_{(3, 85)} = 17.59$, $P < .001$). Bonferroni post hoc tests showed statistically significant differences in students of age 13 years for height (m), weight (kg), and absolute $VO_{2\ Peak}$ compared to all older age groups ($P < .001$), with students of 13 years being shorter, lighter, and having lower absolute measured $VO_{2\ Peak}$. There were no differences in these parameters found among ages 14 to 16 years. Regarding BMI, students of 16 years of age had statistically higher BMI compared to students of 13 years of age ($P < .001$), but no other differences were found among the other age groups. There were no statistically significant differences among age groups in %BF, relative measured $VO_{2\ Peak}$, 1-MRW times, and PACER laps. Figures 2.2 and 2.3 display the linear relationship between measured $VO_{2\ Peak}$ and field test performance from the 1-MRW and PACER, respectively. Table 2.2 displays the Pearson product-moment correlations between measured $VO_{2\ Peak}$ and cardiorespiratory performance parameters, %BF, and BMI. The Pearson correlations between field performance measures, 1-MRW and PACER, with measured $VO_{2\ Peak}$ were considered strong ($P < .001$). Additionally, %BF had a strong correlation with measured $VO_{2\ Peak}$ ($P < .001$); however, BMI did not display a statistically significant relationship with measured $VO_{2\ Peak}$ ($P = .087$).

Table 2.3 displays the results of the cross-validation analysis for the 1-MRW and PACER prediction models. There was a mean overestimation of measured $VO_{2\ Peak}$ across all examined models. Mean differences between estimated and measured $VO_{2\ Peak}$ ranged from 2.84 $ml \cdot kg^{-1} \cdot min^{-1}$ using the Mile-PEQ to 4.95 $ml \cdot kg^{-1} \cdot min^{-1}$ using the Linear PACER model. The multiple correlations (*R*) were considered strong across all models with correlations ranging from $R_{YY'} = 0.74$ to 0.78. The SEE ranged from 5.11 $ml \cdot kg^{-}$

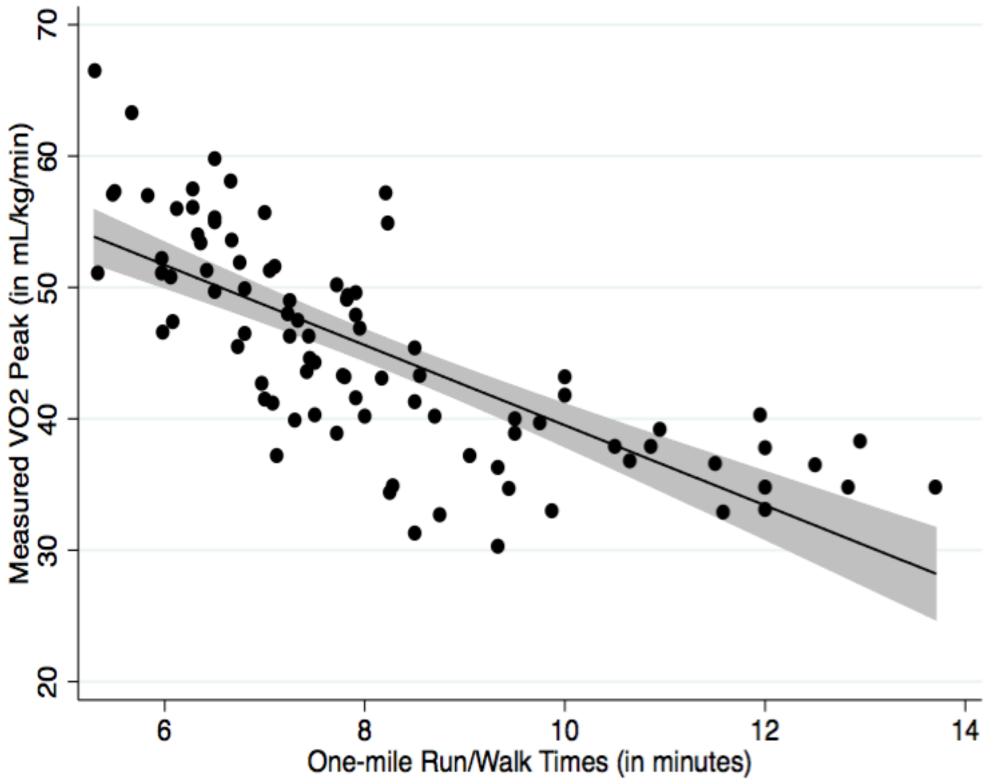

Figure 2.2. Scatterplot and line of best fit showing the linear relationship between measured $VO_{2\ \mathrm{Peak}}$ and 1-MRW times.

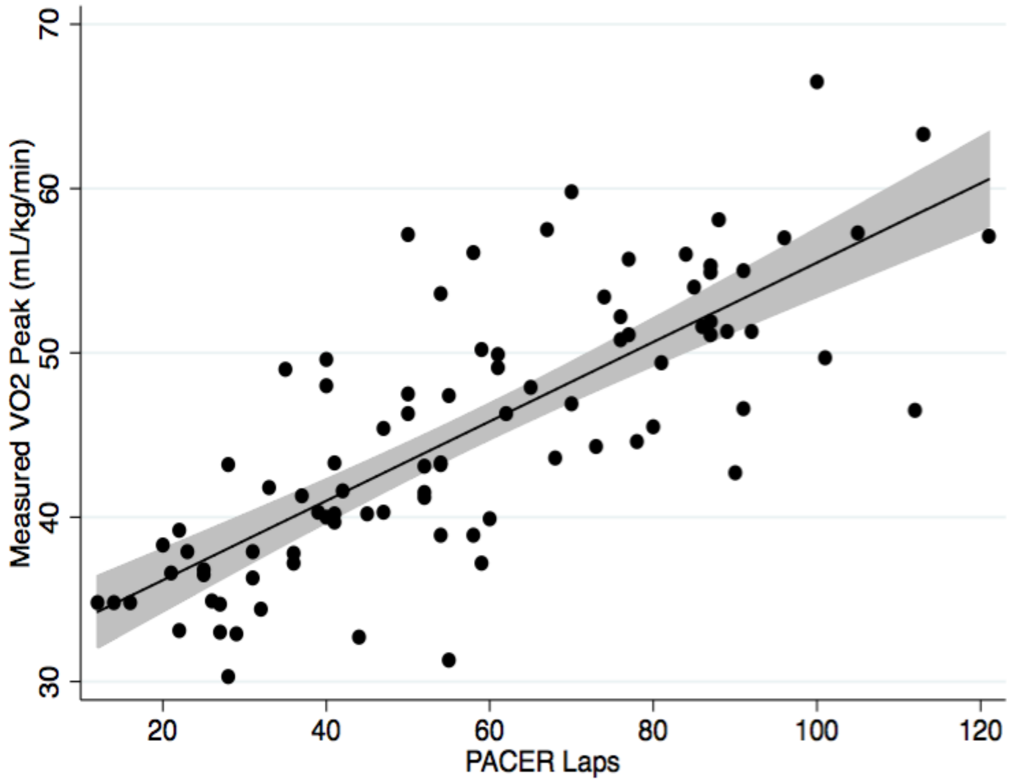

Figure 2.3. Scatterplot and line of best fit showing the linear relationship between measured $VO_{2\ Peak}$ and PACER laps.

Table 2.2. Pearson correlations among cardiorespiratory endurance and body composition parameters.

		$VO_{2\,Peak}$ $(ml \cdot kg^{-1} \cdot min^{-1})$	1-MRW[1] (minutes)	PACER[2] (Laps)	%BF[3] (percentage)	BMI[4] $(kg \cdot m^{-2})$
$VO_{2\,Peak}$ $(ml \cdot kg^{-1} \cdot min^{-1})$	Total					
	Girls	1				
	Boys					
1-MRW[1] (minutes)	Total	**-0.73**[†]				
	Girls	**-0.62**[†]	1			
	Boys	**-0.72**[†]				
PACER[2] (Laps)	Total	**0.77**[†]	**-0.81**[†]			
	Girls	**0.70**[†]	**-0.77**[†]	1		
	Boys	**0.67**[†]	**-0.78**[†]			
%BF[3] (percentage)	Total	**-0.60**[†]	**0.67**[†]	**-0.57**[†]		
	Girls	-0.24	**0.46**[†]	**-0.38**[*]	1	
	Boys	**-0.53**[†]	**0.63**[†]	**-0.41**[†]		
BMI[4] $(kg \cdot m^{-2})$	Total	-0.18	**0.29**[•]	-0.12	**0.49**[†]	
	Girls	-0.09	**0.37**[*]	-0.19	**0.77**[†]	1
	Boys	**-0.28**[*]	**0.32**[*]	-0.14	**0.45**[†]	

[1] 1-MRW stands for the 1-mile Run/Walk test
[2] PACER stands for the Progressive Aerobic Cardiovascular Endurance Run
[3] %BF stands for Percent Body Fat estimated from two-site skin-folds thickness
[4] BMI stands for Body Mass Index
[†] denotes statistical significance, $P < .001$
[*] denotes statistical significance, $P < .05$

Table 2.3. Cross-validation of the aerobic capacity models against measured $VO_{2\ Peak}$.

	t-statistic	Mean Difference[6] (ml·kg^{-1}·min^{-1})	p-value	$R_{YY'}$	Total Error (ml·kg^{-1}·min^{-1})
Cureton (1995)[1]	6.53	3.60	< .001	**0.77**†	6.33
New PACER (2014)[2]	6.75	4.22	< .001	**0.77**†	7.26
Mile-PEQ[3]	5.13	2.84	< .001	**0.76**†	5.95
Mahar Linear (2011)[4]	9.02	4.95	< .001	**0.78**†	7.16
Mahar Quadratic (2011)[5]	5.73	3.31	< .001	**0.74**†	6.38

[1] for Cureton model refer to Equation 1
[2] for New PACER refer to Equation 4
[3] Mile-PEQ stands for $VO_{2\ Peak}$ predicted from the Mile-PACER equivalent score
[4] for Mahar Linear (2011) refer to Equation 2
[5] for Mahar Quadratic refer to Equation 3
[6] mean differences based on (Estimated − Measured $VO_{2\ Peak}$)
† denotes statistical significance, $P < .001$

$^{1}\cdot min^{-1}$ using the Linear PACER to 5.49 $ml\cdot kg^{-1}\cdot min^{-1}$ using the Quadratic PACER. All models were considered to have acceptable prediction error.

Figures 2.4 through 2.8 present the residual versus fitted plots to examine the distribution of the residuals (error) across the range of estimated $VO_{2\ Peak}$ values. Only the Linear PACER model showed a statistically significant correlation between residuals and estimated $VO_{2\ Peak}$ values with a correlation coefficient of $r = -.25$ ($P = .020$). All other models displayed a non-significant distribution (random distribution) of the residuals across the range of estimated $VO_{2\ Peak}$ values, confirming the assumption of homoscedasticity. Table 2.4 shows the CR agreement in the classification of estimated and measured $VO_{2\ Peak}$ into Healthy Fitness Zones. Figure 2.9 depicts the distribution of Healthy Fitness Zone classification for each 1-MRW and PACER model. CR agreement was considered moderate among models with agreement percentages ranging from 72% to 79%.

Discussion

The purpose of this study was to cross-validate various 1-MRW and PACER aerobic capacity prediction models against measured $VO_{2\ Peak}$ in adolescents and to examine CR validity of the models using current FITNESSGRAM standards. This was the first study comparing the New PACER prediction model to older PACER models in addition to the Cureton 1-MRW model in the ability to accurately estimate $VO_{2\ Peak}$ using a sample of adolescent school youth aged 13 to 16 years. The results support previous findings regarding the accuracy of these models to estimate $VO_{2\ Peak}$ but raise some potential issues when administering these models to certain groups of adolescents.

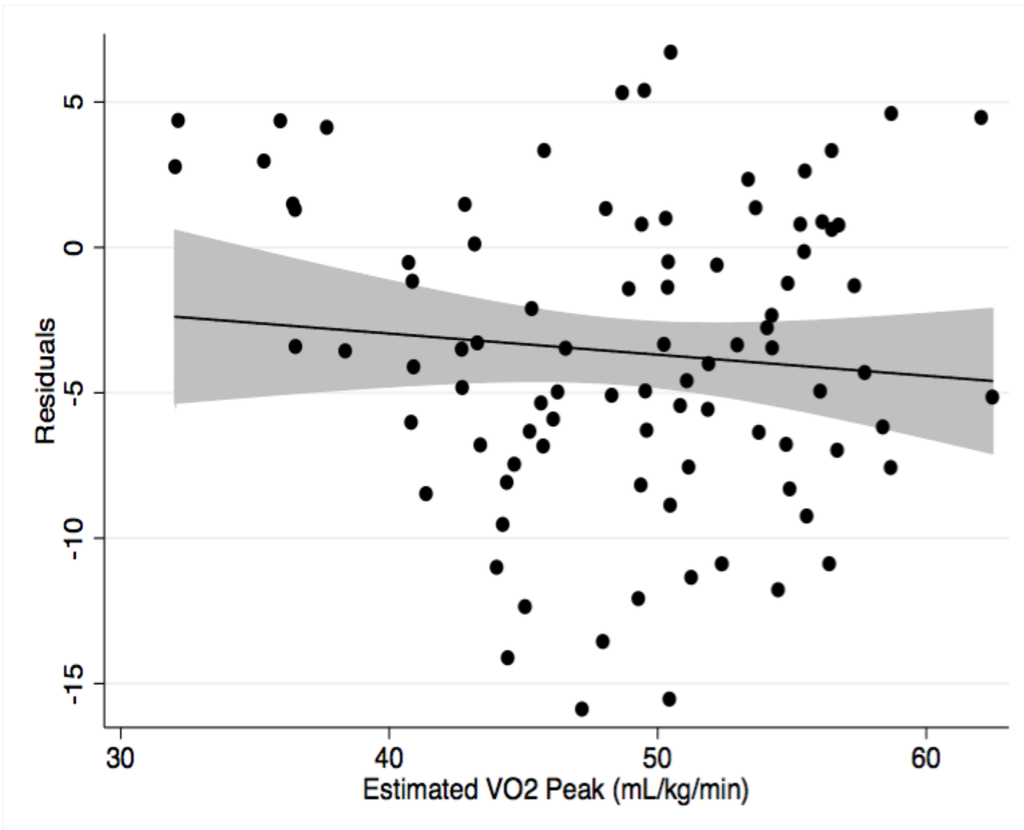

Figure 2.4. Residual against fitted plot with trend line using the Cureton model.

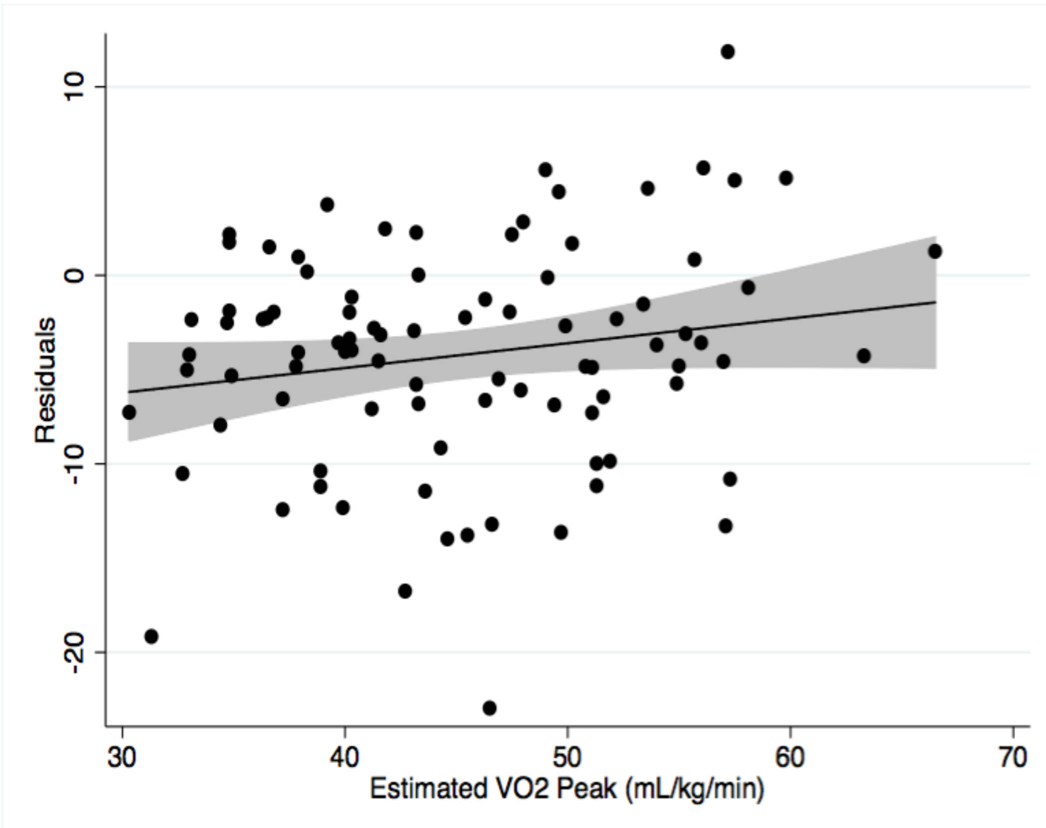

Figure 2.5. Residual against fitted plot with trend line using the New PACER model.

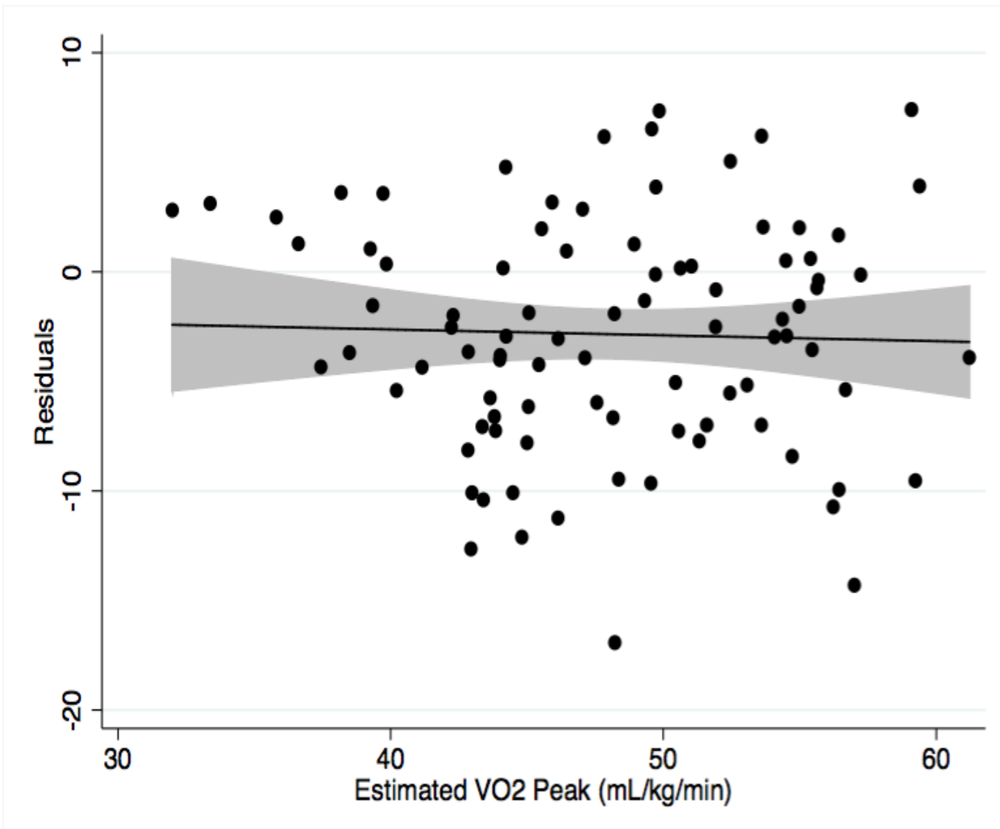

Figure 2.6. Residual against fitted plot with trend line using the Mile-PEQ.

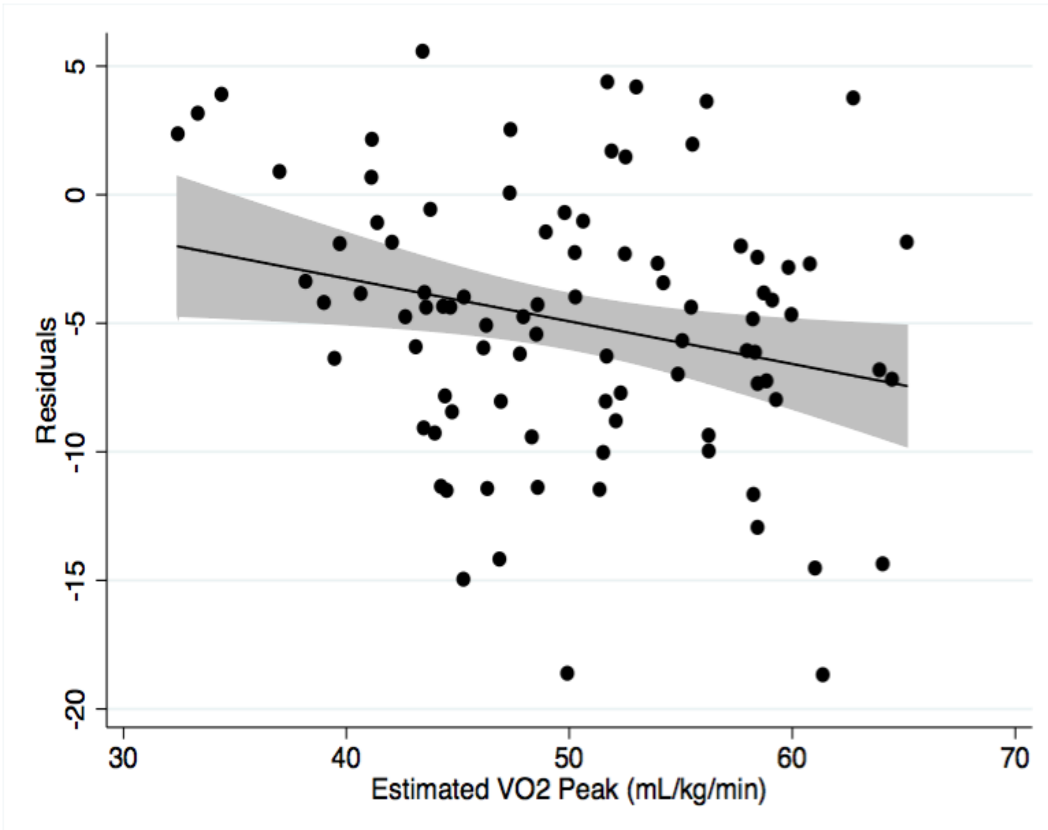

Figure 2.7. Residual against fitted plot with trend line using the Linear PACER model.

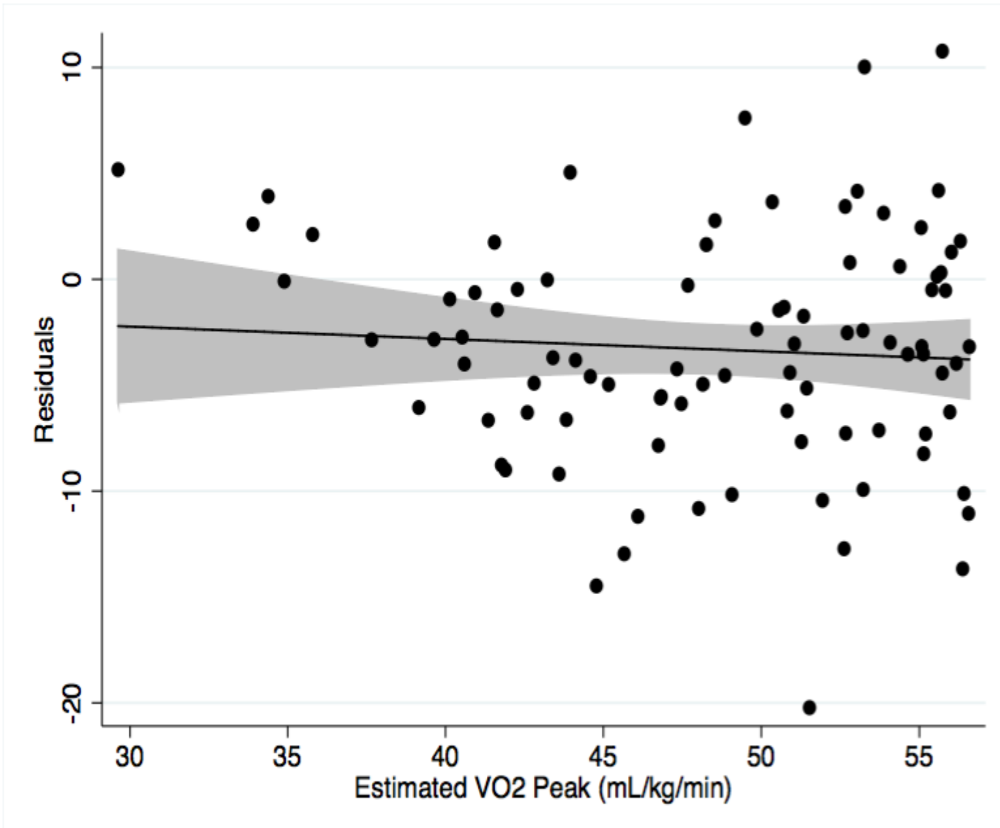

Figure 2.8. Residual against fitted plot with trend line using the Quadratic PACER model.

Table 2.4. CR agreement into FITNESSGRAM's three Healthy Fitness Zone scheme.

	Kq[6]	95% C. I.[7]	*P*-value	*Pa*[8]	ϕ[9]
Cureton (1995)[1]	0.25	0.22, 0.50	< .001	72.2%	0.38
New PACER (2014)[2]	0.49	0.40, 0.54	< .001	79.0%	0.65
Mile-PEQ (2010)[3]	0.28	0.22, 0.52	< .001	73.3%	0.44
Mahar Linear PACER (2011)[4]	0.29	0.08, 0.39	< .001	74.4%	0.46
Mahar Quadratic PACER (2011)[5]	0.29	0.25, 0.39	< .001	74.4%	0.47

[1] for Cureton model refer to Equation 1
[2] for Mahar (2014) refer to Equation 4
[3] Mile-PEQ stands for estimated $VO_{2\ Peak}$ from the Mile-PACER equivalent score
[4] for Mahar Linear (2011) refer to Equation 2
[5] for Mahar Quadratic refer to Equation 3
[6] Kq stands for modified kappa statistic
[7] 95% C.I. stands for the 95% Confidence Interval
[8] *Pa* stands for Proportion of Agreement
[9] ϕ stands for the phi coefficient
[†] denotes statistical significance, $P < .001$

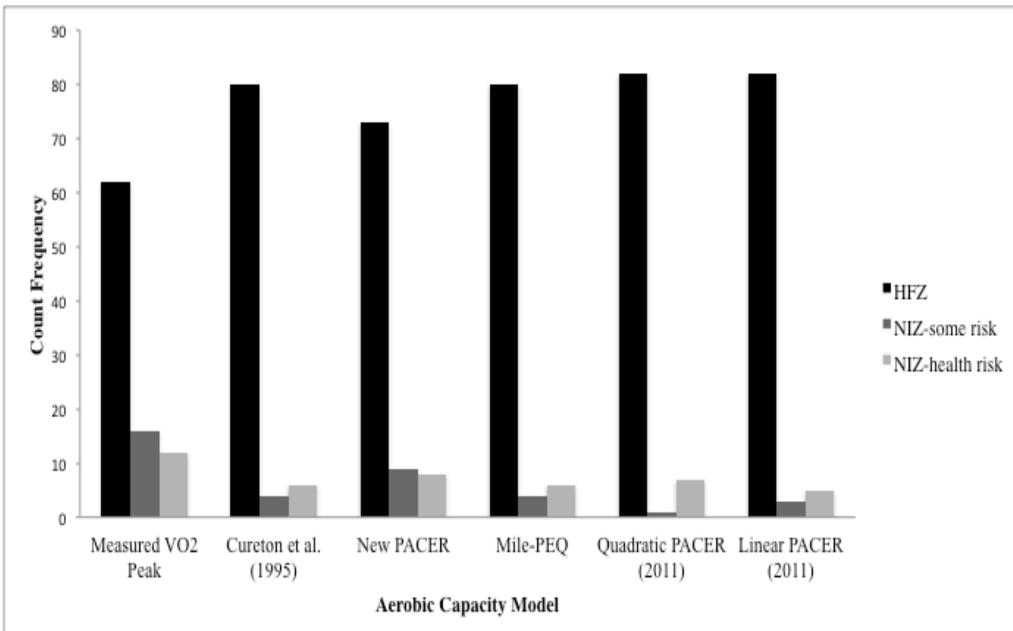

Figure 2.9. Distribution of Healthy Fitness Zone classification for each aerobic capacity model.

Cross-validation analyses of the models support that each model's estimate of $VO_{2\ Peak}$ is strongly correlated with measured $VO_{2\ Peak}$. Indeed, the multiple correlations yielded from this sample were slightly stronger than correlations found in previous research (Boiarskaia et al., 2011; Castro-Pinero et al., 2009; Mahar et al., 2011; Zhu et al., 2010). However, there was a mean overestimation of measured $VO_{2\ Peak}$ across every prediction model examined in this study. This was not expected, as mean relative $VO_{2\ Peak}$, 1-MRW times, and PACER laps from the current study were similar for boys and girls compared to previous work. However, this sample's age range was narrower compared to previous studies examining the validity of these models (Boiarskaia et al., 2011; Cureton et al., 1995; Mahar et al., 2011). Because the models to predict aerobic capacity from 1-MRW and PACER are generalized models, developed from samples using broad age ranges with diverse physical characteristics, the average estimate of $VO_{2\ Peak}$ may not be as accurate when employed to specific subpopulations of youth (adolescents), as was the case in this study.

The PACER models examined in this study all showed strong linear relationships with measured $VO_{2\ Peak}$, with prediction error ranging from 5.95 to 7.26 $ml \cdot kg^{-1} \cdot min^{-1}$. Of all the models examined, the New PACER model developed from Mahar et al. (2014) was the only one not containing a BMI term. The inclusion of BMI may cause problems for teachers because of the need to collect height and weight information. Additionally, if a teacher is including BMI for body composition assessment when using the FITNESSGRAM battery, a student with a higher BMI will essentially be penalized twice: once for body composition assessment and once for cardiorespiratory endurance assessment by lowering his/her estimated $VO_{2\ Peak}$. Double penalization due to high BMI may not be an overall fair assessment protocol because body composition and cardiorespiratory endurance are separate domains of health-related physical fitness (Welk et al., 2011). There also may be potential for misclassification at the individual level if a student has higher BMI due to higher levels of fat-free mass (FFM) rather than higher levels of fat mass (FM).

Although a BMI term may marginally contribute increased predictive ability at the population level, at the individual level there is large variability in the relationship between BMI and cardiorespiratory endurance, especially across age ranges where significant increases in FFM and FM begin to accrue (adolescence; Goran et al., 2004; Malina, Bouchard, and Bar-or, 2004). In this study, mean BMI (20.6 $kg \cdot m^{-2}$ for boys; 20.5 $kg \cdot m^{-2}$ for girls) was similar compared to Boiarskaia et al. (2011; 20.6 $kg \cdot m^{-2}$ for boys; 20.6 $kg \cdot m^{-2}$ for girls), Cureton et al. (1995; 19.5 $kg \cdot m^{-2}$ for boys; 18.8 $kg \cdot m^{-2}$ for girls), and Mahar et al. (2011; 20.4 $kg \cdot m^{-2}$ for boys; 20.5 $kg \cdot m^{-2}$ for girls). However, in this study the correlation between BMI and measured $VO_{2\ Peak}$ was statistically insignificant (r = -0.18. P = 0.087), where in previous research the addition of a BMI predictor significantly added to the predictive accuracy of the generalized models (Cureton et al., 1995; Mahar et al., 2011). The New PACER model was the only model examined in this study that did not use BMI in its prediction algorithm. For the New PACER equation, total error, mean differences between estimated and measured $VO_{2\ Peak}$ (absolute accuracy), and multiple correlations between the model's estimate and the measured value (relative accuracy) were similar compared to the other PACER models examined in this study. Therefore, the performance of the New PACER, a model that does not include BMI in its prediction of aerobic capacity, compared favorably to the other PACER models that included BMI within a norm-referenced framework.

The Cureton 1-MRW model also displayed strong multiple correlations and low prediction error with measured $VO_{2\ Peak}$. Unfortunately, there is currently only one model for the 1-MRW and no valid alternative that can be used for comparison. Correlations between measured $VO_{2\ Peak}$ and 1-MRW times vary greatly across studies due to sample size differences, differences in age ranges, and student experience running the 1-MRW. Zero-order correlation coefficients have ranged from r = -0.54 to -.80 (Buono et al., 1991; Castro-Pinero et al., 2009; Cureton et al., 1995). In smaller homogeneous samples these correlations tend to be stronger compared to larger more heterogeneous samples (Cureton et al., 1995). When developing models to estimate $VO_{2\ Peak}$, previous authors have found it necessary to include demographic predictors such as age and gender, along with BMI, because in most studies 1-MRW alone explained less than 50% of the variance in measured $VO_{2\ Peak}$ (Castro-Pinero et al., 2009; Cureton et al., 1995).

Cureton et al. (1995) developed the 1-MRW algorithm using multiple regression modeling from a sample of 753 individuals aged 8 to 25 years. The model consists of several predictors including 1-MRW times, a quadratic term for 1-MRW, an age x gender interaction term, and BMI. The model accounted for 50% of total variance in measured $VO_{2\ Peak}$ in the original validation sample, 55% of the total variance in the cross-validation sample, and 52% of the variance in measured $VO_{2\ Peak}$ using the total sample. The standard errors of estimates were 4.78 $ml \cdot kg^{-1} \cdot min^{-1}$ in the validation sample, 4.99 $ml \cdot kg^{-1} \cdot min^{-1}$ in the cross-validation sample, and 4.84 $ml \cdot kg^{-1} \cdot min^{-1}$ using the total sample. A limitation of the study by Cureton et al. (1995), however, is that data were collected on three separate samples across a 19-year time frame (1972 to 1991). Therefore, a number of confounders may have influenced the results, including but not limited to inconsistency of laboratory and field testing protocols, different measurement techniques for obtaining measured $VO_{2\ Peak}$, and use of different test administrators. Additionally, the majority of students were laboratory tested using a graded treadmill-walking test, and it was not stated in the manuscript whether the children met criteria that would have indicated they achieved a valid $VO_{2\ Peak}$ measurement. Therefore, the validity of the Cureton model in explaining variance of an individual's measured $VO_{2\ Peak}$ is questionable.

When examining the CR validity evidence among 1-MRW and PACER models, similar agreement values with measured $VO_{2\ Peak}$ compared to previous research were displayed, with percentages of agreement ranging from Pa = 72% to 79%. High levels of CR agreement between estimated and measured $VO_{2\ Peak}$ are important because the CR standards used in FITNESSGRAM reflect health status. Indeed, Lobelo, Pate, Dowda, Liese, and Ruiz (2009) examined the clinical utility of FITNESSGRAM's CR standards and found that the older standards (two Healthy Fitness Zones) were able to discriminate adolescent youth aged 12 to 19 years with and without CVD risk factors, which included hemostatic model assessment of insulin resistance (HOMA-IR), triglycerides, and total cholesterol/HDL ratio. The sample in this study was quite large (N = 1,247); however, the assessment of cardiorespiratory endurance was a submaximal walking treadmill test, which is a less accurate measure of aerobic capacity compared to a maximal treadmill test.

A few studies have examined the newer CR standards, which include classification into a three Healthy Fitness Zone scheme. Mahar et al. (2011) examined the CR validity of various PACER models used to estimate $VO_{2\ Peak}$ and found agreement values with measured $VO_{2\ Peak}$ ranging from 67% to 70% using three Healthy Fitness Zones for classification. The highest agreement values were found using the Quadratic PACER model, which included a BMI term. However, there were no comparisons between these prediction models with Mile-PEQ in the ability to accurately classify students into the Healthy Fitness Zones. Boiarskaia, Boscolo, Zhu, and Mahar (2011) cross-validated the Mile-PEQ method using an independent sample of middle-school students (N = 135) and found agreement values ranging from 73% to 75% between Mile-PEQ estimated $VO_{2\ Peak}$ and measured $VO_{2\ Peak.}$ These agreement values were similar to Mahar's Linear PACER (Pa = 75%) and Mahar's Quadratic PACER (Pa = 73%) models, suggesting similar accuracy among models in middle school students in classifying students into Healthy Fitness Zones based on PACER performance. The percentages of agreement yielded from the current study were nearly identical compared to the aforementioned studies that examined PACER model CR agreement with measured $VO_{2\ Peak}$ in adolescents.

This is the first study to examine CR agreement between both 1-MRW and PACER prediction models estimates with measured $VO_{2\ Peak}$ using current CR standards in 13 to 16 year old adolescents. Recent evidence from NHANES data suggests that 42.2% of American youth aged 12 to 15 years have inadequate levels of cardiorespiratory endurance using FITNESSGRAM age and gender specific standards (Gahche et al., 2014). The distribution of students classified into the NI subzones in the current study was approximately 33%. Therefore, the sample had a slightly more favorable distribution of students into the HFZ compared to national data. The results indicate that the agreement with measured $VO_{2\ Peak}$ for both 1-MRW and PACER prediction models are quite similar; however, CR agreement into the three Healthy Fitness Zone scheme is still classified as moderate due to percentage of agreement values below 80% across all models. Several potential sources of error may affect classification accuracy.

The Cureton, Mile-PEQ, and Mahar Linear and Quadratic models inherently favor students with lower BMI, who are older, and are male due to the direction of the parameters in the model. For these models, if 1-MRW or PACER performance is the same among students, students who are lighter, older, or are male will have a higher estimated $VO_{2\ Peak}$ compared to students who are heavier, younger, or are female. At the individual level these predictors may be a potential source of misclassification. Indeed, Castro-Pinero et al. (2009) found no differences between normal and overweight children aged 8 to 17 years on measured $VO_{2\ Peak}$, but did find differences between BMI groups on $VO_{2\ Peak}$ estimated from the Cureton model. Several other confounders that may have accounted for additional unexplained variance in measured $VO_{2\ Peak}$ and influence classification accuracy include genetics, maturation, and running economy.

The New PACER Model (Mahar et al., 2014) addresses the problem of having a BMI term in a model by incorporating only PACER laps and age as predictors of $VO_{2\ Peak}$. The model displayed relatively higher prediction error compared to the other PACER models examined in this study. This is possibly due to the use of only two predictors (PACER Laps and Age) compared to using four predictors in the other models. However, the marginal decrease in predictive accuracy did not affect CR agreement with measured $VO_{2\ Peak}$ using the New PACER model. On the contrary, the CR validity of the New PACER model was quite high, displaying a percentage of agreement of Pa = 79%, stronger than any other 1-MRW or PACER model. Therefore, the CR validity of the New

PACER model is strong given the evidence from this study. However, the prediction of $VO_{2\ Peak}$ from 1-MRW still needs to be addressed. Additional models that do not contain BMI must be developed from 1-MRW in order to provide an alternative to Cureton et al. (1995).

Limitations in this study should be considered before generalizations can be made. First, the sample consisted of adolescent youth aged 13 to 16 years; therefore, results cannot be generalized to younger or older age groups. Second, the 1-MRW was administered outside for all students on different tracks either on school grounds or at the University's facilities. This was done for convenience and to reduce subject burden; however, different weather conditions and tracks may have affected the results obtained in this study. Lastly, different students participating in this study performed the PACER test in different settings (in physical education class, before sport practices, alone at the University facilities). Again, this was done for convenience and to reduce subject burden. However, the lack of a standard setting for each field measure may have affected the results. However, measures were taken to minimize differences between subject testing protocols, as explained previously. Therefore, the effects of differences in field test environment on physical performance were assumed to be minimal.

This is the first study to examine the external validity of both 1-MRW and PACER prediction models in a sample of adolescent youth aged 13 to 16 years. The results support the conclusion of strong norm-referenced validity (strong R, low prediction error) and moderate CR agreement with measured $VO_{2\ Peak}$ into Healthy Fitness Zones. This was also the first study to examine the external validity of the New PACER model (Mahar et al., 2014). The evidence from this study support the use of the

New PACER model in adolescent school-age children as it compared favorably to other PACER models and displayed the highest CR agreement with measured $VO_{2\ Peak}$. The 1-MRW Cureton model did show relatively low prediction error with measured $VO_{2\ Peak}$ and moderate CR agreement into Healthy Fitness Zones; however, the inclusion of BMI as a predictor variable in the model does manifest potential issues for physical education teachers. This issue was addressed in the following study. In conclusion, both 1-MRW and PACER prediction models provided similar predictive accuracy and CR agreement with measured $VO_{2\ Peak}$ into Healthy Fitness Zones. The New PACER compared favorably to the other models examined in this study and shows promise for great utility in physical education settings.

CHAPTER 3

DEVELOPMENT OF AN AEROBIC CAPACITY PREDICTION MODEL FROM 1-MILE RUN/WALK PERFORMANCE IN ADOLESCENTS

Introduction

Fitness assessment is an important component of many school physical education programs in the United States. Effective assessment informs students of their current fitness status and may serve as a health warning sign to students who may have fitness levels that are associated with increased health risk. The FITNESSGRAM youth fitness assessment program is widely used in schools to facilitate the collection and reporting of fitness data (Welk et al., 2011). FITNESSGRAM uses criterion-referenced (CR) standards to evaluate students on whether they are at a level of fitness needed for good health. A domain of health-related fitness that is strongly related to health outcomes is cardiorespiratory endurance (Blair et al., 1989; Eisenmann et al., 2004; Ogden et al., 2010; Ortega et al., 2008).

The 1-mile run/walk test (1-MRW) or the Progressive Aerobic Cardiovascular Endurance Run (PACER) is recommended to assess cardiorespiratory endurance (Meredith & Welk, 2010). These cardiorespiratory field tests estimate aerobic capacity, operationally defined by $VO_{2\ Peak}$, which is the maximal amount of oxygen one can take in, transport, and utilize during exercise. The new CR standards for cardiorespiratory endurance were developed using a large nationally representative National Health and Nutrition Examination Survey (NHANES) database by relating $VO_{2\ Peak}$ with the metabolic syndrome criteria (Laurson et al., 2011; Welk et al., 2011). Age and gender-specific cut-off scores were developed using receiver operating characteristic curve (ROC) analyses that yielded three distinct Healthy Fitness Zones. These Healthy Fitness Zones are currently used for classification purposes in physical education/fitness assessment settings to inform students of their current fitness status and how it relates to their health (Zhu et al., 2011).

Although the PACER test is the recommended cardiorespiratory endurance test in FITNESSGRAM, the 1-MRW is still used extensively in physical education settings. $VO_{2\ Peak}$ is estimated from 1-MRW times using the Cureton et al. (1995) model (see equation 1). This model has established validity and utility in the youth population; however, the BMI coefficient within the model manifests a potential issue in fitness assessment by potentially overestimating $VO_{2\ Peak}$ in lighter children and underestimating $VO_{2\ Peak}$ in heavier children.

The Cureton et al. (1995) model makes sense from a practical standpoint in that children carrying more weight will tend to have lower relative measured $VO_{2\ Peak}$. However, a child who has a higher BMI will essentially be penalized twice when using FITNESSGRAM: once for the body composition assessment (if using BMI) and once again during cardiorespiratory endurance assessment by the lowering of his or her estimated $VO_{2\ Peak}$ when employing the Cureton et al. model. This may be perceived as an unfair protocol because body composition and cardiorespiratory endurance are separate domains of health-related physical fitness (Welk et al., 2011).

Physical educators often use BMI as the test of body composition in youth assessment, even though two-site skinfold thickness (SKF) is FITNESSGRAM's recommended test. One of SKF's limitations is that it is time consuming, so administering this assessment to a large physical education class is not efficient. It may also increase student discomfort due to the skin pinching involved (Hannon, Ratliffe, & Willams, 2006), and for reliable and valid measurement, training is needed (Opplinger, Clark, & Kuta, 1992; Shaw, 1986). Due to these limitations of SKF, BMI is often preferred for body composition assessment because it is easier to administer and calculate.

However, BMI does contain potential limitations, which include that it does not take into account the relative contribution of fat-free mass (FFM) and fat mass (FM) to body weight, nor does BMI specify the degree of central adiposity which has been more strongly linked to health outcomes in adults and children than total body adiposity which is predicted with BMI (Caprio, Hyman, & Limb, 1995; Himes et al., 2009; Morrison et al., 1999; Savva, Tornaritis, & Savva, 2000). Therefore, there is potential for underestimation of true aerobic capacity when employing the Cureton et al. (1995) model if a child has a higher BMI largely due to higher amounts of FFM relative to FM, which may facilitate higher cardiorespiratory endurance (faster 1-MRW times). Conversely, the model has potential to overestimate aerobic capacity in students with relatively lower BMI, but higher FM relative to FFM, contributing to lower cardiorespiratory endurance (slower 1-MRW times). Additionally, the BMI term within the Cureton et al. model necessitates the collection of height and weight information by the physical educator.

These measures may not be readily available due to time limitations and large class sizes.

Because of the aforementioned issues of incorporating a BMI term into an aerobic capacity prediction equation, alternative models should be developed to circumvent these limitations. Therefore, the purpose of this study was to develop alternative aerobic capacity prediction models from 1-MRW performance from a sample of adolescent youth aged 13 to 16 years. It was hypothesized that the validated models from this study would yield similar predictive accuracy compared to the Cureton et al. (1995) model. The similar predictive accuracy of the new models would not significantly affect the criterion-referenced (CR) agreement between measured and estimated $VO_{2\ Peak}$ in classifying students into Healthy Fitness Zones. The results from this study will yield models with great utility in physical education settings due to the elimination of the practical issues related to incorporating a BMI predictor variable in an aerobic capacity model.

Methods

Participants and Setting

The participants of this study were 90 adolescent youth (38 girls, 52 boys) recruited from middle and high schools located in a metropolitan area in the southwestern United States. Data collection took place in both laboratory and field settings. The laboratory setting included having the students report to the HPRL before or after school hours for $VO_{2\ Peak}$ testing. The field settings included either testing students on school grounds or testing them at the University gymnasium or outdoor track. Written assent was obtained from the students and written consent was obtained from the parents prior to data collection. The University Institutional Review Board (IRB) approved the protocols used in this study.

Study Procedures

Day 1 consisted of having students complete the 1-MRW on school grounds during their physical education class or at the University's facilities. On Day 2, students reported to the HPRL either after school hours or on weekends to complete $VO_{2\ Peak}$ testing. $VO_{2\ Peak}$ testing occurred no less than 48 hours following the 1-MRW and no more than 3 weeks following the 1-MRW. Students were instructed to wear comfortable gym attire and to not have eaten within 2 hours prior to testing. Prior to the $VO_{2\ Peak}$ testing, height and weight were collected via a portable stadiometer (Seca 213; Chino, CA, USA) and medical scale (Tanita HD-314; Arlington Heights, IL, USA). Percent body fat was estimated via two-site skinfold thickness on the triceps and calf measured three times in a rotating order using a Lange skinfold caliper (Ann Arbor, MI, USA). Three consecutive measurements that were within ± 2 mm were averaged per site. The Slaughter formula was used to estimate %BF from the two measurements (Slaughter et al., 1988). Students were then familiarized with the $VO_{2\ Peak}$ testing protocol and performed the graded maximal exercise test as described earlier. In order to maintain testing consistency, all data were collected by the same trained graduate student in both laboratory and field settings.

Data Analysis

Multiple linear and polynomial regression modeling were used to develop two prediction models from 1-MRW times. Measured $VO_{2\ Peak}$ ($ml \cdot kg^{-1} \cdot min^{-1}$) was used as the dependent variable. Hierarchical (block-wise entry) multiple linear and polynomial regression was employed to develop prediction models for measured $VO_{2\ Peak}$. The first entry block consisted of the 1-MRW predictor variable. The second block consisted of an age, gender, or an age x gender interaction term. The third block consisted of a 1-MRW^2 predictor. Finally, the fourth block consisted of BMI or %BF to test if the addition of a body composition term could significantly contribute to predictive accuracy of the model. Inclusion of a predictor into the model was determined by significant changes in R^2, adjusted R^2, in addition to the statistical significance of the coefficient.

This statistical methodology yielded two prediction models, one linear model and one quadratic model. Multicollinearity was assessed using the Variance Inflation Factor (V.I.F.) with a cut-off of 10. Testing the assumptions on the regression models included examining the normality of the residuals via Normal Probability plots, plots of residual versus fitted data to identify outliers and assess the assumption of constant variance (homoscedasticity), and examination of a scatterplot matrix to determine linearity between the dependent variable and independent predictors. In addition to examining the aforementioned plots, homoscedasticity was tested statistically using the Breusch-Pagan test. Influential cases were identified using Cook's Distance with a cut-off for case removal set at ≥ 1 a priori.

The equations developed on the total sample were cross-validated using k-fold cross-validation. Statistically significant terms from the developed models using the total sample were tested using a five-iteration process. Iterations from k-fold cross-validation involved splitting the total sample into approximately five equal subsamples. On four of the five subsamples (training data) a model was fitted and then cross-validated using the remaining subsample (testing data). The measure of fit for each developed model on the testing data was assessed using the standard error of the estimate (SEE). This process was repeated four more times for a total of five iterations. Therefore, each case in the dataset was incorporated into the training subsample exactly k − 1 or 4 times and the testing sub-sample exactly once. If the coefficients and average SEE from k-fold cross-validation were similar to the coefficients and SEE from the developed models using the total sample, the validity of the linear and quadratic models was supported. The multiple correlations (*R*), coefficients of determination (R^2), and SEE were then compared between the final two 1-MRW prediction models developed in this study and the Cureton et al. (1995) model.

CR validity was examined by comparing each estimated $VO_{2\ Peak}$ score's classification with measured $VO_{2\ Peak}$ score's classification into the three Healthy Fitness Zone scheme. CR agreement was analyzed using weighted kappa statistics (*Kq*), percentage of agreement (*Pa*), and a phi coefficient (ϕ). Proportion of Agreement into Healthy Fitness Zones was considered poor if below 60%, moderate if 60% to 80%, and excellent if above 80% (Hartmann, 1977). Statistical analyses were conducted using STATA v12.0 (College Station, TX, USA) statistical software.

Results

The descriptive, anthropometric, and cardiorespiratory endurance score data for the total sample and within each gender group are presented in Table 2.1 (See Chapter 2). times. Two separate models were developed using the total sample. Figure 3.1 shows the curvilinear relationship between measured $VO_{2\ Peak}$ and 1-MRW times. Two separate

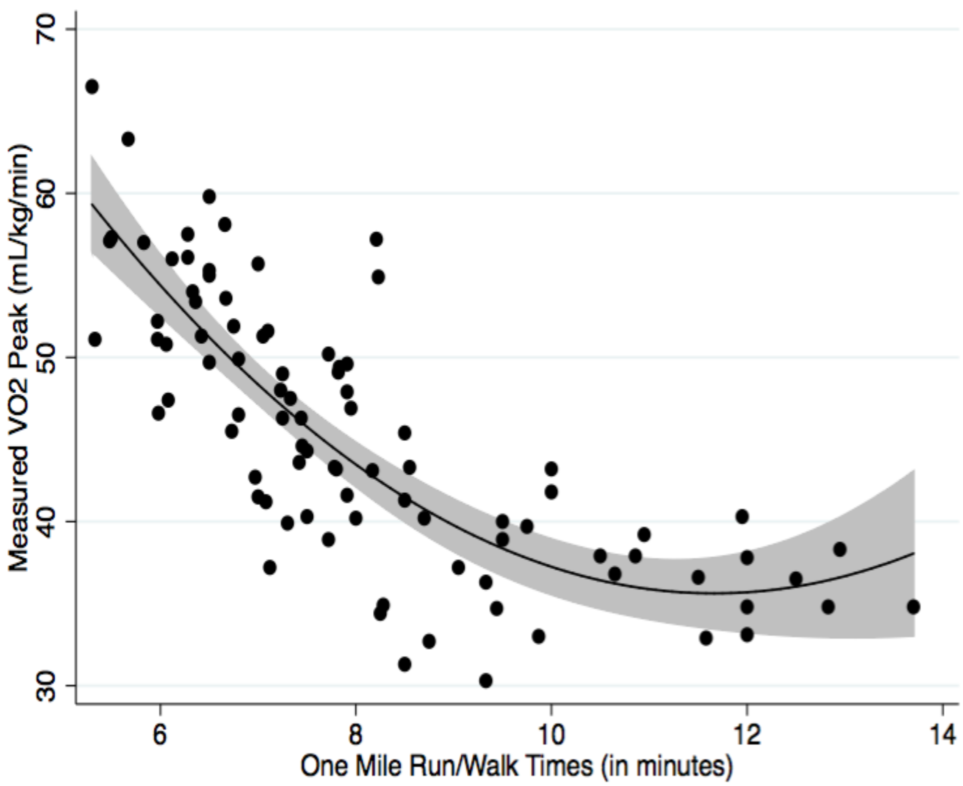

Figure 3.1. Scatterplot and best fit line showing the curvilinear relationship between measured $VO_{2\ Peak}$ and 1-MRW times.

models were developed using the total sample. Table 3.1 displays the model fit parameters using the block-wise entry method. Model parameters are presented in Table 3.2 for the linear model and Table 3.3 for the quadratic model. The linear model explained 60% of the variance in measured $VO_{2\ Peak}$ with a SEE of 5.21 $ml \cdot kg^{-1} \cdot min^{-1}$, and the quadratic Model explained 67% of the variance in measured $VO_{2\ Peak}$ with a SEE of 4.72 $ml \cdot kg^{-1} \cdot min^{-1}$. In contrast, the Cureton et al. (1995) model explained 59% ($R = .76$) of variance in measured $VO_{2\ Peak}$ with a SEE of 5.20 $ml \cdot kg^{-1} \cdot min^{-1}$. The new 1-MRW linear and quadratic models are represented by the following equations, respectively:

$$VO_{2\ Peak} = -2.47(1\text{-MRW}) + 0.29(\text{Gender} \times \text{Age}) + 62.85 \quad [6]$$

$$VO_{2\ Peak} = -12.07(1\text{-MRW}) + 0.52(1\text{-MRW})^2 + 0.22(\text{Gender} \times \text{Age}) + 104.93 \quad [7]$$

where 0 = girl and 1 = boy for Gender,

Age = chronological age in years,

and 1-MRW = 1-mile run/walk time in minutes.

Figures 3.2 and 3.3 show the Normal Probability plots for the regression residuals from the linear and quadratic models, respectively. Both models showed evidence for a normal distribution of their residuals; however, the quadratic model displayed a slightly greater deviation from normality compared to the linear model. Figures 3.4 and 3.5 depict the residual versus fitted plots examining the distribution of residuals across the range of estimated aerobic capacity levels. The linear model showed some pattern of non-linearity. The quadratic model displayed a random distribution of the residuals across the range of estimated $VO_{2\ Peak}$, confirming the assumption of homoscedasticity.

Table 3.1. Model development employing hierarchical block-wise entry.

Model	R	R^2	Adjusted R^2	SEE ($mL \cdot kg^{-1} \cdot min^{-1}$)
Block 1[b]	0.73	0.54	0.54	5.51
Block 1 + Block 2[c]	0.77	0.60	0.59	5.21
Block 1 to Block 3[d]	0.82	0.67	0.66	4.72
Block 1 to Block 4[e]	0.83	0.68	0.67	4.69
Block 1 to Block 3 + Block 5[f]	0.83	0.68	0.67	4.69

Note: Dependent variable is measured $VO_{2\ Peak}$
b Block 1 predictor is 1-MRW
c Block 2 predictor is Age x Gender
d Block 3 predictor is $1\text{-}MRW^2$
e Block 4 predictor is BMI
f Block 5 predictor is %BF

Table 3.2. Parameter estimates for linear model.

	β -Coefficient	Standard Error	p-value	V.I.F[2]
Constant	62.85	3.16	< .001	
1-MRW (min.)[1]	-2.47	0.33	< .001	1.28
Age x Gender	0.29	0.09	< .001	1.28

[1]1-MRW stands for the 1-mile Run/Walk test
[2]V.I.F. stands for Variance Inflation Factor

Table 3.3 Parameter estimates for quadratic model.

	β -Coefficient	Standard Error	p-value	V.I.F[3]
Constant	104.93	9.81	< .001	
1-MRW (min.)[1]	-12.07	2.16	< .001	74.39*
Age x Gender	0.22	0.08	.008	1.34
1-MRW^2 (min.^{2})[1]	0.52	0.12	< .001	72.14*

[1]1-MRW stands for the 1-mile Run/Walk test
* High V.I.F. due to use of polynomial regression modeling.

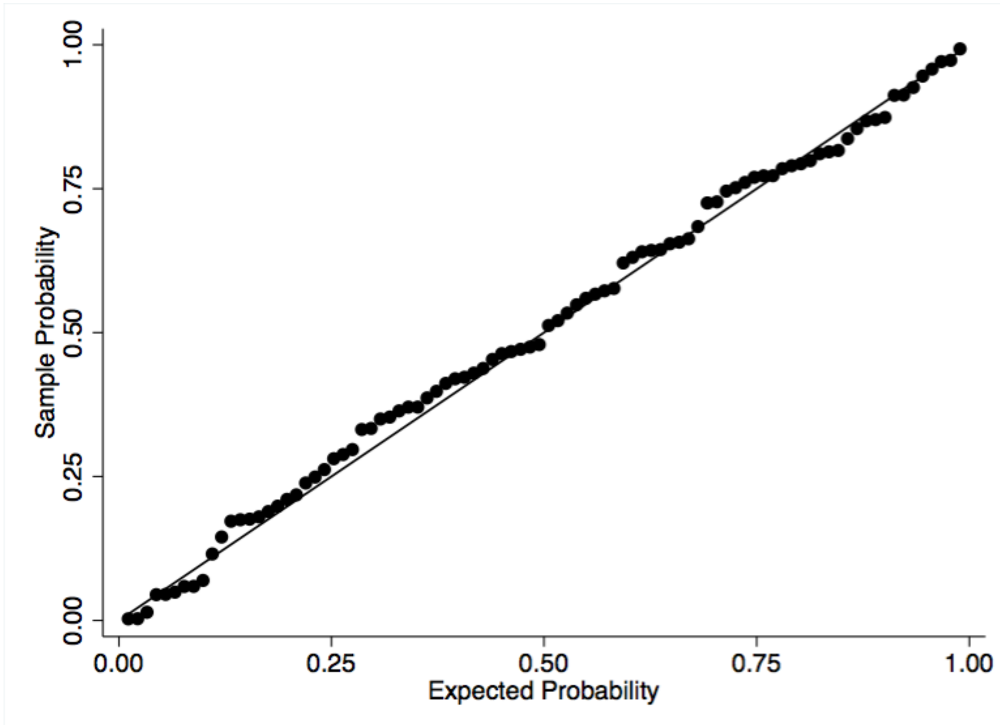

Figure 3.2. Normal probability plot of residuals for the linear model.

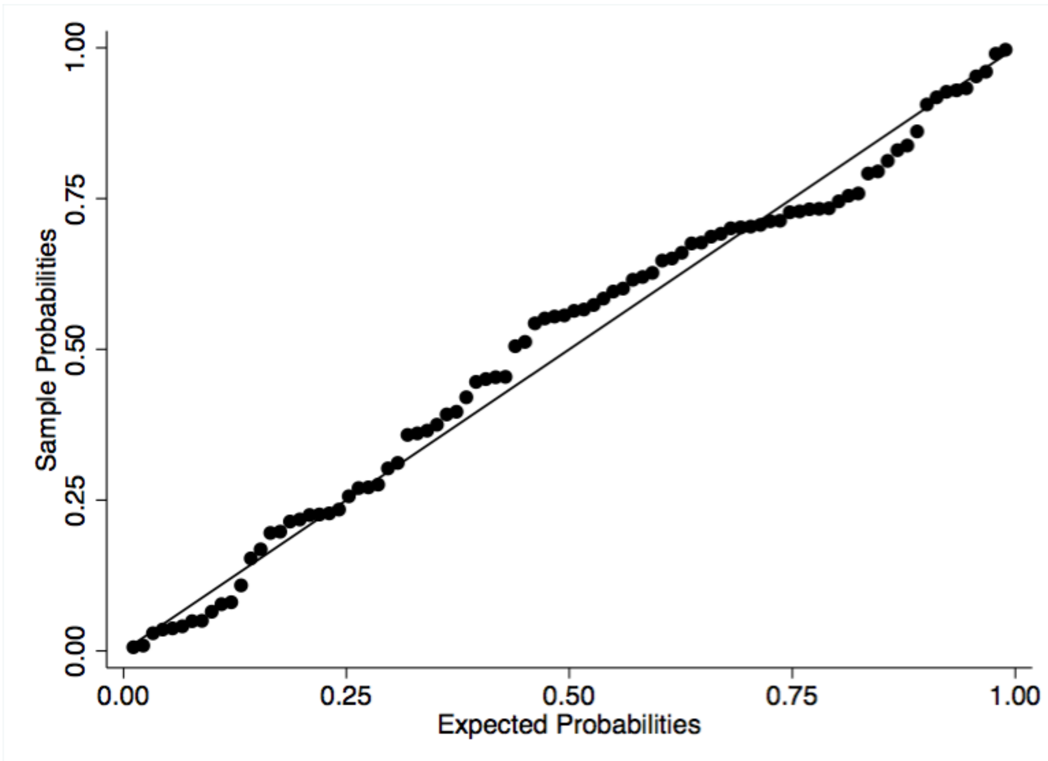

Figure 3.3. Normal probability plot of residuals for the quadratic model.

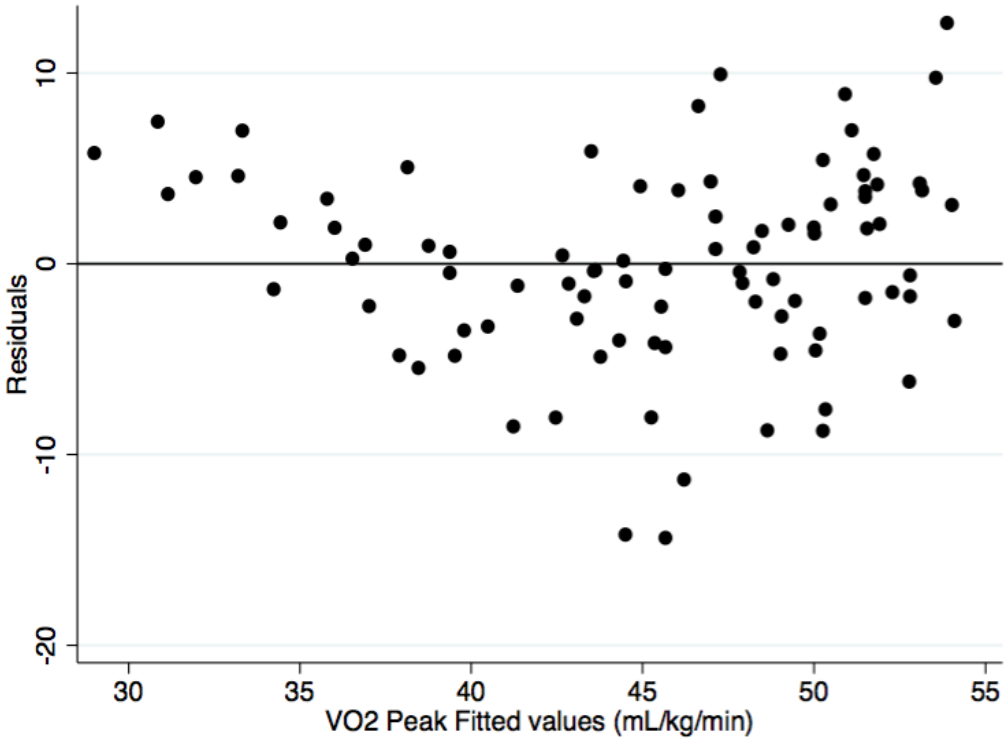

Figure 3.4. Residual against fitted plot using the new 1-MRW linear model.

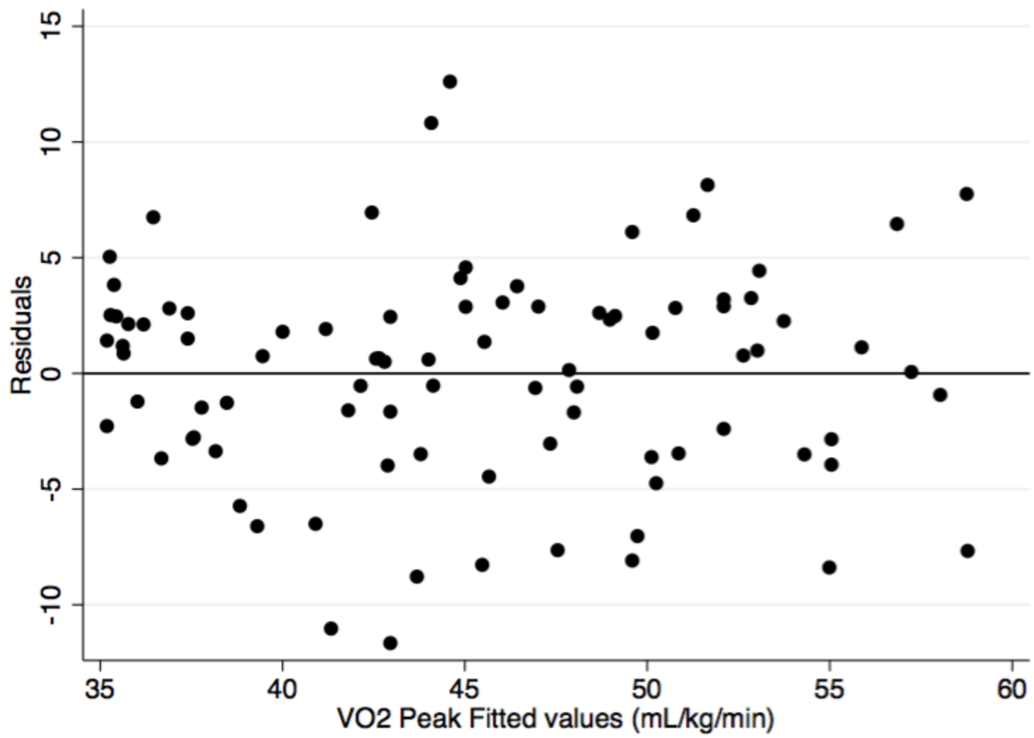

Figure 3.5. Residual against fitted plot using the new 1-MRW quadratic model.

Results from each Breuch-Pagan test for each model yielded statistically nonsignificant chi-square statistics: (χ^2 = 0.79, P = 0.37) for the linear model and (χ^2 = 1.03, P = 0.31) for the quadratic model. These results suggest that the assumption of homoscedasticity was not violated for either the Linear or Quadratic model.

Results from the k-fold cross-validation are presented in Table 3.4 for the linear model and Table 3.5 for the quadratic model. Because the iteration coefficients for each model were not statistically different compared to the coefficients of the developed models, and the average 5-fold cross-validation SEE was similar to the linear and quadratic model's SEE, the validity of the two models developed using the total sample was supported. Model CR agreement with measured $VO_{2\ Peak}$ is presented in Table 3.6, and the distribution of classification across Healthy Fitness Zones is visually depicted in Figure 3.6. Proportion of agreement with measured $VO_{2\ Peak}$ was considered moderate among the new models and Cureton et al. (1995). Agreement values ranged from 70% to 73% into the three Healthy Fitness Zone scheme.

Discussion

The purpose of this study was to develop and cross-validate two aerobic capacity prediction models from 1-MRW performance. The two new models compared favorably to the older Cureton et al. (1995) model, displaying good predictive accuracy and moderate CR agreement with measured $VO_{2\ Peak}$ into Healthy Fitness Zones. The 1-MRW is a popular test of cardiorespiratory endurance in school-aged children. FITNESGRAM gives the physical educator the option of choosing among three cardiorespiratory

Table 3.4. K-fold cross-validation of the 1-MRW linear model.

	Cross-validation SEE ($ml \cdot kg^{-1} \cdot min^{-1}$)	Coefficients from Model (95% C.I.)		
		Constant	1-MRW Time	Age x Gender
Iteration 1	3.93	65.06	-2.68	0.24
		(57.72, 72.40)	(-3.43, -1.93)	(0.04, 0.45)
Iteration 2	5.29	61.78	-2.35	0.25
		(54.73, 68.84)	(-3.08, -1.63)	(0.05, 0.45)
Iteration 3	6.20	62.80	-2.40	0.30
		(56.18, 69.43)	(-3.10, -1.72)	(0.12, 0.49)
Iteration 4	4.06	61.80	-2.49	0.33
		(55.22, 88.30)	(-3.28, -1.70)	(0.13, 0.54)
Iteration 5	6.63	61.80	-2.37	0.34
		(55.22, 68.38)	(-3.07, -1.67)	(0.16, 0.52)
Average	5.22	-	-	-

Table 3.5. K-fold cross-validation of the 1-MRW quadratic model.

	Cross-validation SEE ($ml \cdot kg^{-1} \cdot min^{-1}$)	Coefficients from Model (95% C.I.)			
		Constant	1-MRW Time	Age x Gender	1-MRW $Time^2$
Iteration 1	4.69	104.51	-12.16	0.28	0.53
		(83.09, 125.94)	(-16.84, -7.47)	(0.09, 0.47)	(0.28, 0.79)
Iteration 2	5.70	112.94	-13.54	0.16	0.58
		(92.54, 133.34)	(-18.05, -9.04)	(-.0.02, .32)	(0.34, 0.83)
Iteration 3	4.38	106.69	-12.44	0.18	0.54
		(81.10, 132.28)	(-18.20, -6.70)	(.00, .35)	(0.22, 0.85)
Iteration 4	4.82	100.96	-11.23	0.24	0.48
		(79.14, 122.80)	(-15.99, -6.46)	(0.07, 0.43)	(0.22, 0.73)
Iteration 5	4.19	99.73	-11.08	0.26	0.48
		(78.07, 121. 40)	(-5.82, -6.35)	(0.06, 0.45)	(0.22, 0.73)
Average	4.76	-	-	-	-

Table 3.6. CR agreement with measured $VO_{2\ Peak}$ into Healthy Fitness Zones.

	Kq[4]	95% C.I.[5]	*P*-value	*Pa*[6]	ϕ[7]
Linear Model[1]	0.38	0.35, 0.46	< .001	0.73	0.57
Quadratic Model[2]	0.34	0.24, 0.36	< .001	0.70	0.54
Cureton et al. (1995)[3]	0.25	0.22, 0.49	< .001	0.72	0.37

[1] for linear model refer to Equation 5
[2] for quadratic model refer to Equation 6
[3] for Cureton model refer to Equation 1
[4] Kq stands for modified kappa statistic
[5] 95% C.I. stands for the 95% Confidence Interval
[6] Pa stands for Proportion of Agreement
[7] ϕ stands for the phi coefficient

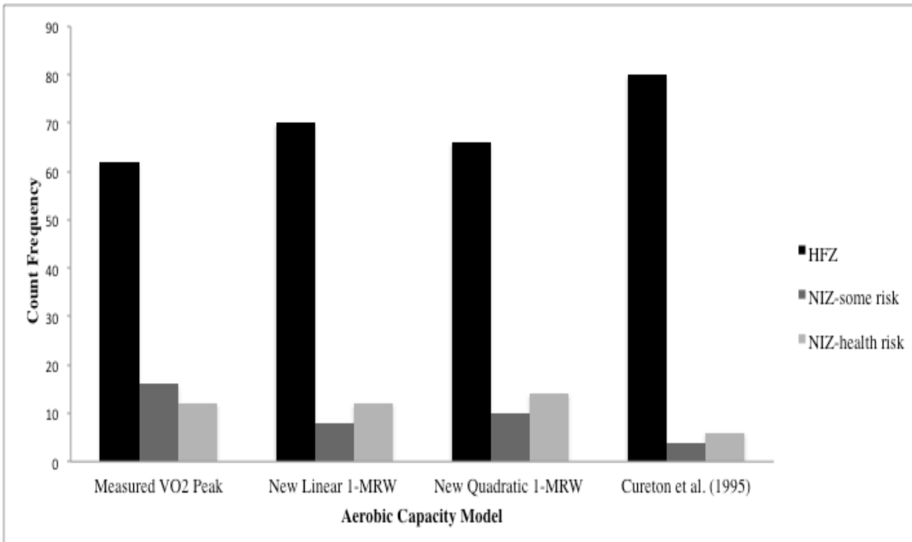

Figure 3.6. Distribution of Healthy Fitness Zone Classification for each 1-MRW aerobic capacity model.

endurance tests that are used to predict aerobic capacity: the 1-MRW, PACER, and 1-Mile Walk. In order to interpret performance of these field tests in terms of cardiorespiratory function and health and to compare performances among the three field tests, FITNESSGRAM employs prediction models to estimate $VO_{2\ Peak}$ from one of the three aforementioned field measures.

Although the PACER is the recommended (default) assessment when using FITNESSGRAM (Meredith & Welk, 2010), the 1-MRW is still widely employed as a fitness test to assess cardiorespiratory endurance. The 1-MRW is time efficient, uses limited resources, and can be administered to large class sizes, accounting for the popularity of this test among physical educators worldwide (Australian Council for Health, Physical Education, and Recreation, 1996; Cooper Institute for Aerobics Research, 2004; The President's Council on Physical Fitness and Sports, 2007). Despite its popularity, the generalized prediction model used to estimate $VO_{2\ Peak}$ from 1-MRW has been without a valid alternative for approximately 20 years. The Cureton et al. (1995) model, currently used by FITNESSGRAM to estimate $VO_{2\ Peak}$ from 1-MRW, includes a BMI term that necessitates the collection of height and weight information, double penalizes students with higher BMI when also assessing body composition, and may lead to misclassification of students who have higher BMI due to higher levels of FFM relative to FM.

The two models developed from this study compared favorably with the older Cureton et al. (1995) model; however, similarities and contrasts among the models must be discussed. In this sample, the zero-order correlation between measured $VO_{2\ Peak}$, and 1-MRW was $r = -.73$. This correlation is stronger than that found by Castro-Pinero et al.

(2009; $r = -.59$) and Cureton et al. (1995; $r = -.60$), but identical to that found in Bono, Roby, Micale, Sallis, and Shepard (1991; $r = -.73$) using a sample size similar to this study. The differences found in this study compared to that of Cureton et al. in developing aerobic capacity models from 1-MRW may be due to the use of an older, more specific age range (13 to 16 years) compared to the broader (8 to 25 years) age range used in the Cureton et al. study. The 1-MRW may not be as valid of a test in younger students compared to older youth in predicting aerobic capacity because younger students may have trouble developing an appropriate pace during the assessment in addition to poor motivation, low tolerance to physical discomfort, and underdeveloped motor coordination needed for running (Cureton, Baumgartner, & McManis, 1991; Krahenbuhl, Pangrazi, Burkett, Schneider, & Petersen, 1977). Therefore, because this study's sample exclusively consisted of older adolescent youth, the correlations between 1-MRW and measured VO_{2Peak} may have been stronger compared to studies where samples were comprised of a broader age range.

A squared term for the 1-MRW was determined to be a significant predictor of measured $VO_{2\ Peak}$; therefore, the inclusion of a squared term in a separate model was warranted. This is consistent with what was found in the study by Cureton et al. (1995) as the addition of the quadratic term into the model significantly increased its predictive accuracy. At 1-MRW times slower than the inflection point at approximately 11.6 minutes (using first derivative approximation), there was a slightly positive linear relationship between 1-MRW and measured $VO_{2\ Peak}$; across 1-MRW times faster than 11.6 minutes, there was a clear negative linear relationship (see Figure 3.1). The positive linear relationship between 1-MRW and measured $VO_{2\ Peak}$ at 1-MRW scores greater (slower) than the point of inflection may be attributable to the behaviors that manifest in adolescents who display longer 1-MRW times, such as lack of motivation, incorrect pacing, or decreased tolerance to prolonged physical exertion. Therefore, for children whose 1-MRW times are especially slow, performance on the field test becomes less of a function of oxygen consumption and more attributable to the aforementioned behavioral confounders not controlled for in this study. This limitation of incorporating a quadratic term for 1-MRW in cardiorespiratory endurance assessment may lead to potential misclassification within a CR framework because it is a nonlocal polynomial model that, at the individual level, may award students who do not display the behavioral confounders that are displayed in other students while performing the 1-MRW.

The 1-MRW linear and quadratic models also contained an Age x Gender interaction term. This interaction term suggests there was an average higher measured $VO_{2\ Peak}$ between two cohorts of boys separated in age by 1 year. However, in girls, older age cohorts did not result in an average higher measured $VO_{2\ Peak}$. This is consistent with the findings of Eisenmann et al. (2011), who showed slight increases in $VO_{2\ Peak}$ across the ages of 12 to 18 years in boys, but very slight decreases in average $VO_{2\ Peak}$ in girls across the same age range. Differences in relative oxygen consumption capabilities between the age cohorts and genders may be attributable to the increases in FFM that accrue during physical development in boys and the increase in FM that accrues in girls (Malina, Bouchard, & Bar-or, 2004) and decreases in social acceptance of physical activity in adolescent girls. Therefore, the interaction between age and gender in a regression model specific to adolescents aged 13 to 16 years is theoretically and statistically defensible. However, this finding does not generalize to students younger than 12 years of age, as the relationships between age, gender, and $VO_{2\ Peak}$ may be different in preadolescents and younger children as found in the more generalized models (i.e., New PACER model).

When comparing the new models to the Cureton et al. (1995) model, the new models contain fewer predictor variables. Although mean measured $VO_{2\ Peak}$, 1-MRW times, and BMI were similar in this sample compared to aforementioned studies, this study produced regression models using only 1-MRW times, age x gender, and 1-MRW^2 predictors. The inclusion of either BMI or %BF did not statistically increase the predictive accuracy of either the linear or quadratic model. BMI did not display a statistically significant zero-order correlation with measured $VO_{2\ Peak}$ ($r = -0.18$, $P = 0.08$); however, the correlation between %BF and measured $VO_{2\ Peak}$ was strong ($r = -.60$, $P < .001$). Despite this, the additions of either body composition predictor into the regression models did not explain any additional unique variance in measured $VO_{2\ Peak}$ after the other predictors were entered into the model. This may have been due to the strong zero-order correlations between 1-MRW times and measured $VO_{2\ Peak}$ in adolescents. Therefore, the addition of a body composition predictor was not needed to yield a model with good predictive accuracy after accounting for the other predictors in the model. In samples using younger age groups this may not be the case as the correlation between 1-MRW times and measured $VO_{2\ Peak}$ may be significantly weaker for reasons explained previously. Therefore several additional predictor variables may be needed to yield acceptable explanatory power. Because the two models developed in this study had nearly the same predictive accuracy and CR validity with measured $VO_{2\ Peak}$ as compared to the Cureton et al. (1995) model, the evidence suggests that there is no need to use a body composition predictor in estimating $VO_{2\ Peak}$ in adolescents.

Each model developed in this study displayed strong linear relationships with measured $VO_{2\ Peak}$, acceptable prediction error, and moderate CR agreement with measured $VO_{2\ Peak}$ into the Healthy Fitness Zones. Indeed, the fit parameters, prediction error, and agreement values were similar among the two new models and the Cureton et al. (1995) model. Therefore, in adolescents, the inclusion of BMI as a predictor variable in the model did not increase predictive accuracy or CR validity. Additionally, results from k-fold cross-validation showed stable coefficients for each model across all iterations and an average error (RMSE) that was nearly identical to the developed models RMSE. This supports the predictive accuracy of the models developed using the total sample. Despite good predictive accuracy, the CR agreement with measured $VO_{2\ Peak}$ into Healthy Fitness Zones was only classified as moderate. Mahar et al. (2011) found similar agreement values into a three Healthy Fitness Zone scheme when developing PACER models to estimate $VO_{2\ Peak}$. Zhu et al. (2010) also found agreement percentages ranging from 69% to 75% when classifying PACER and 1-MRW estimated $VO_{2\ Peak}$ into three categories (low, medium, high) with measured $VO_{2\ Peak}$. Therefore, given the evidence from previous studies, it should be reasonably expected that models' CR agreement into three Healthy Fitness Zones should range from anywhere between 70% and 75% with measured $VO_{2\ Peak}$. With the exception of Study 1 of this dissertation using the New PACER, agreement percentages above this range have not been detected.

Limitations of this study should be considered before generalizations can be made. The sample consisted of students between the ages of 13 to 16 years; therefore, the results cannot be generalized to younger or older age groups. The 1-MRW was administered on an outside track for all students; therefore, slight changes in the weather (i.e., temperature differences, etc.) may have affected testing performance. Additionally, some students performed the 1-MRW with other students, and some performed the 1-MRW independently. This was done for convenience and to reduce subject burden in the context of scheduling; however, it may have affected the results by influencing confounding variables not controlled for in this study, such as intent and motivation.

In conclusion, the linear and quadratic models developed in this study compared favorably to Cureton et al. (1995) model using a sample of adolescent youth, displaying good predictive accuracy and moderate CR agreement with measured $VO_{2\ Peak}$ into Healthy Fitness Zones. The use of models that do not contain a BMI term eliminates the issues of obtaining height and weight information, double penalization, and potential for misclassification that the inclusion of BMI manifests at the individual level within an aerobic capacity prediction model. Future research needs to externally validate the new models using a large diverse sample so it can be generalized to larger populations of school-aged children. The results from this study support that aerobic capacity prediction models that do not contain a body composition term can provide a fair and accurate prediction of aerobic capacity in adolescents.

CHAPTER 4

SUMMARY OF FINDINGS

Cross-validation of Mahar PACER (2014)

The New PACER model compared favorably to previous models that are used to estimate $VO_{2\ Peak}$ from PACER performance. PACER models examined in Study 1 included the Mahar Linear (2011), Mahar Quadratic (2011), Mile-PEQ, and the New PACER model developed from Mahar et al. (2014). These select models were examined because they were the most recently developed and have displayed the strongest predictive accuracy and CR agreement with measured $VO_{2\ Peak}$ (Boiarskaia et al., 2011; Mahar et al., 2011; Zhu et al., 2010). All PACER models displayed strong linear relationships with measured $VO_{2\ Peak}$. However, there was a mean overestimation of measured $VO_{2\ Peak}$ across all models examined in this study. This specific finding may have been attributable to the use of a limited sample size and a narrow age-range (adolescents). All of the PACER models examined were developed in previous studies using large sample sizes consisting of school-aged children with broad age ranges and diverse physical characteristics. Zhu et al. (2010) also found that various models' estimates of $VO_{2\ Peak}$ were systematically different compared to measured values. Despite this specific finding, the New PACER model displayed similar mean differences with measured $VO_{2\ Peak}$ compared to the other models; therefore, its performance using this sample of adolescents was no different from the Mile-PEQ, or the Mahar Linear and Quadratic models (2014) within a norm-referenced framework.

Within a criterion-referenced framework, CR agreement was considered moderate among all models with agreement percentages ranging from 72% to 79% with measured $VO_{2\ Peak}$. These agreement percentages are similar to the percentages found in previous research examining the CR validity of these PACER models (Boiarskaia et al., 2011; Mahar et al., 2011; Zhu et al., 2010). Even though the New Mahar model only incorporates a PACER score and age term in the model, it had the strongest CR agreement with measured $VO_{2\ Peak}$ into the Healthy Fitness Zones. This is encouraging because the New PACER model does not contain a BMI term, which may cause problems for physical educators. The New PACER circumvents the issues of including a BMI term, which include the need to collect height and weight information, double penalizing children who have higher BMI when also assessing body composition along side cardiorespiratory endurance, and the potential for misclassification for students who have higher BMI but elevated cardiorespiratory endurance levels (high PACER scores). Therefore, the results from this study indicate that the New PACER model has great utility for physical educators in estimating $VO_{2\ Peak}$ from the PACER in adolescents.

Cross-validation of Cureton et al. (1995)

The Cureton et al. (1995) model is currently the prediction model used in FITNESSGRAM to estimate $VO_{2\ Peak}$ from 1-MRW in school-aged children. Results from the first study suggest that this model displayed strong predictive accuracy and moderate CR agreement with measured $VO_{2\ Peak}$ into the three Healthy Fitness Zone scheme. The model did systematically overestimate measured $VO_{2\ Peak}$ in this sample. The amount of overestimation was similar in magnitude to that found for the PACER models. The Cureton et al. model inherently tends to decrease estimated aerobic capacity in lighter children, when all other parameters in the model are held constant, because of the BMI term. This may lead to potential misclassification in children with low BMI but relatively low cardiorespiratory endurance (slower 1-MRW times) and in children with higher BMI but relatively higher levels of cardiorespiratory endurance (faster 1-MRW times). Additionally, employing the Cureton model requires collecting height and weight information. These measures may not be readily available for physical educators because of large class sizes and time constraints. There is currently no alternative to the Cureton et al. (1995) model for estimation of $VO_{2\ Peak}$ from 1-MRW times.

Development of New 1-MRW Models

Two new aerobic capacity prediction models were developed from 1-MRW. Both models displayed strong predictive accuracy and moderate CR agreement with measured $VO_{2\ Peak}$, comparing favorably to the Cureton et al. (1995) model. When developing the models, the addition of a BMI or a %BF predictor did not statistically increase predictive accuracy. Because the relationship between 1-MRW times and measured $VO_{2\ Peak}$ was curvilinear, the quadratic model displayed a stronger multiple correlation and lower SEE compared to the linear model. However, the addition of a squared term within an aerobic capacity prediction model may manifest interpretation problems when assessing cardiorespiratory endurance. Past the inflection point of a 1-MRW time of approximately 11 minutes, the linear relationship between 1-MRW and measured $VO_{2\ Peak}$ slightly positively increases. Therefore, students with slower 1-MRW times will have higher estimated $VO_{2\ Peak}$ compared to students who have faster 1-MRW times who perform at or above approximately 11.6 minutes. This may be a source of misclassification within a criterion-referenced framework and possibly does not represent 1-MRW performance as a function of $VO_{2\ Peak}$, but rather the behaviors displayed by students with slow 1-MRW times. Indeed, the Linear Model displayed a slightly stronger CR agreement with measured $VO_{2\ Peak}$ compared to the Quadratic Model. This is an important consideration when choosing to employ a 1-MRW model to estimate $VO_{2\ Peak}$ in physical education settings. Despite this, however, both of the new models without a BMI term compared favorably to the Cureton et al. model in predictive accuracy and CR validity. As is the case with the PACER models, the inclusion of a BMI term can present problems for teachers that include the need to collect height and weight information and difficulty interpreting scores in heavier students who have higher levels of cardiorespiratory endurance (faster 1-MRW times) compared to lighter students. The results from this study provide evidence that models that do not contain a body composition term can have utility in estimating $VO_{2\ Peak}$ from 1-MRW in school-settings by eliminating the limitations that BMI potentially manifests in an aerobic capacity model.

Limitations

This dissertation does include some limitations that must be considered before any generalizations can be made. The sample was limited in the number of subjects not only to develop the models in Study 2 but also for model cross-validation. Although robust statistical techniques were employed using k-fold cross-validation, the models developed in Study 2 must be externally validated to assure they have not overfit the data and can therefore be generalized to greater populations of youth. The students consisted of 13- to 16-year old youth; therefore, the models may not be generalized to younger or older age groups. The 1-MRW was administered outdoors for all students; therefore, weather may have affected the results. Although each student was assessed in fair weather conditions, temperature fluctuations could not be controlled. Additionally, some students performed the field assessments (1-MRW and PACER) with other students, and some students performed the assessments independently. Although this was done for convenience, it may also have affected field test performance by influencing intent, motivation, and other psychometric variables that were not controlled for in this study.

Future Research Directions

Of primary importance is to externally validate the newly developed 1-MRW models using large and diverse samples of school-aged children. Doing so will provide further insights regarding the predictive accuracy of the new models developed in this study. Future research may also need to test other predictor variables not examined in this study to determine if they significantly contribute to the predictive accuracy of aerobic capacity models. Additionally, the 1-MRW is the alternative to PACER for aerobic capacity assessment using FITNESSGRAM, using shorter length variations of the 1-MRW (such as ¼ mile run or ½ mile run) may prove as a more practical assessment of $VO_{2\ Peak}$, especially in younger children.

Conclusions

This dissertation provides solid evidence for the predictive accuracy of a recently developed PACER model from Mahar et al. (2014) and provides two alternative models to Cureton et al. (1995) for estimation of $VO_{2\ Peak}$ from 1-MRW in adolescents. Evidence from this dissertation suggests that prediction models that do not contain a BMI term can be accurate in estimating $VO_{2\ Peak}$ from field test performance and in classifying students into Healthy Fitness Zones. The elimination a body composition term may marginally decrease prediction accuracy (SEE) at the population level in adolescents, but may attenuate the potential for misclassification for individual students who have higher BMIs, but elevated cardiorespiratory endurance and make assessment of cardiorespiratory endurance more fair when employing the FITNESSGRAM battery. Also, the physical educator does not have the need to collect height and weight information when using the New PACER or the New 1-MRW models. Assessment of height and weight may not be feasible in physical education classes with large class sizes and time limitations. The use of a prediction model without a BMI term or a %BF term, given the results from this study, is defensible statistically because the addition of a body composition term when developing the new 1-MRW models did not significantly increase their predictive ability. Additionally, the New PACER model, which does not have a BMI term, had the strongest CR agreement with measured $VO_{2\ Peak}$ compared to all other PACER models, suggesting that a body composition term within an aerobic capacity prediction model may not be needed for acceptable CR validity.

Fitness assessment is an important feature of physical education, and this dissertation advanced the understanding of the validity of various prediction models to estimate $VO_{2\ Peak}$ in adolescent school-aged children. It also provided new information by developing two new aerobic capacity prediction models that do not contain a BMI term, but can still provide acceptable levels of predictive accuracy and CR classification agreement with measured $VO_{2\ Peak}$. This new information will hopefully advance the field of fitness assessment in school-aged children and provide physical educators or practitioners working with children feasible, fair, and valid options to assess cardiorespiratory endurance in school settings.

APPENDIX A

INSTITUTIONAL REVIEW BOARD APPROVAL LETTER

INSTITUTIONAL REVIEW BOARD
THE UNIVERSITY OF UTAH

75 South 2000 East Salt Lake City, UT 84112 | 801.581.3655 | IRB@utah.edu

IRB: IRB_00063008

PI: Ryan Burns

Title: Development and Cross-Validation of an Alternative Aerobic Capacity Prediction Model for Adolescent Youth.

This New Study Application qualifies for an expedited review by a designated University of Utah IRB member as described in 45 CFR 46.110 and 21 CFR 56.110. The research involves one or more activities in Category X (published in 63 FR 60364-60367). The designated IRB member has reviewed and approved your study as a Minimal risk study on 4/16/2013. The approval is effective as of 4/16/2013. Federal regulations and University of Utah IRB policy require this research protocol to be re-reviewed and re-approved prior to the expiration date, as determined by the designated IRB member.

Your study will expire on 4/15/2015.
Any changes to this study must be submitted to the IRB prior to initiation via an amendment form.

APPROVED DOCUMENTS

Parental Permission Forms
Parental Permission Form

Assent Forms
Assent Form

Grant Application
Grant Application

Literature Cited/References
FG2 References

Recruitment Materials, Advertisements, etc.
Recruitment Flyer

Click IRB_00063008 to view the application and access the approved documents.

Please take a moment to complete our customer service survey. We appreciate your opinions and feedback.

APPENDIX B

PARENTAL CONSENT

Consent Document

BACKGROUND

Your child is being asked to take part in a research study to develop a more valid aerobic capacity prediction equation for the FITNESSGRAM fitness assessment program used in physical education curricula. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you will allow your child to take part in this study. The purpose of this study is to develop and then cross-validate a prediction equation for aerobic fitness for adolescent youth (8^{th}-10^{th} grade). Depending on the results, the new formula will replace the current formula that estimates aerobic capacity in physical education classes that use FITNESSGRAM'S assessments.

STUDY PROCEDURE

It will take your child approximately 3-4 testing sessions to complete the study. Your child will report to the University of Utah Human Performance Research Laboratory to complete a treadmill VO2 Peak test. This involves running on a treadmill at a progressively increased intensity until your child voluntarily stops. During this test your child's heart rate will be monitored and pulmonary gas exchange will be measured by having your child breath into a tube while running on a treadmill. They will also report to the University Human Performance Research Laboratory to get their body composition measured via the BOD POD, skinfolds thickness assessment, and Body Mass Index (height and weight). The BOD POD requires your child to sit quietly inside a closed chamber for approximately 5 minutes. Skinfolds thickness assessment requires us to lightly pinch a fold of skin on your child's upper arm and lower leg to get a measurement. At a later date they will then be tested using the one-mile run and PACER aerobic fitness tests. A trained graduate student of your child's same gender and/or a trained faculty member will administer all tests from the Department of Exercise and Sport Science.

RISKS

The risks of this study are minimal. Your child might encounter some minor risks while participating in some fitness tests. The lab VO2 Peak test requires exercising (running) on a treadmill until exhaustion while monitoring heart rate and pulmonary gas exchange data (breathing data). The one-mile run and PACER tests also require cardiovascular exercise to exhaustion. Therefore, there will be the normal physical discomfort involved similar to that of engaging in maximal vigorous exercise in their physical education class or in some athletic/sport programs. Your child might also have emotional stress when we measure their body fat %. However, all body fat measurements will occur in an isolated and controlled environment, **administered by a trained graduate student of**

your child's same gender (males will measure males, females will measure females).

BENEFITS
Benefits include learning about your child's physical fitness levels and how it relates to their current health status. We hope the information we get from this study may help develop a greater understanding of the factors that contribute to aerobic fitness, which is related to cardiovascular health in pediatric and adult populations.

CONFIDENTIALITY
Your child's data will be kept confidential. Data and records will be stored in a locked filing cabinet or on a password protected computer located in the researcher's work space. Only the researcher and faculty sponsor will have access to this information. Results of the study will be published in a journal, but no names or identifying information will be included in the publication.

PERSON TO CONTACT
If you have questions, complaints or concerns about this study, you can contact Ryan Burns at 801-695-5693 or at ryan.d.burns@utah.edu. You may also contact Dr. James C. Hannon at 801-581-7646. If you feel your child has been harmed as a result of participation, please call Ryan Burns or Dr. James C. Hannon who may be reached during 10:00 am – 5:00 pm Monday-Friday.

Institutional Review Board: Contact the Institutional Review Board (IRB) if you have questions regarding your child's rights as a research participant. Also, contact the IRB if you have questions, complaints or concerns which you do not feel you can discuss with the investigator. The University of Utah IRB may be reached by phone at (801) 581-3655 or by e-mail at irb@hsc.utah.edu.

Research Participant Advocate: You may also contact the Research Participant Advocate (RPA) by phone at (801) 581-3803 or by email at participant.advocate@hsc.utah.edu.

VOLUNTARY PARTICIPATION
It is up to you to decide whether to allow your child to take part in this study. Refusal to allow your child to participate or the decision to withdraw your child from this research will involve no penalty or loss of benefits to which your child is otherwise entitled. This will not affect your or your child's relationship with the investigator.

COSTS AND COMPENSATION TO PARTICIPANTS
Your child will receive $25 as compensation for participation in this study after completion of all the tests.

CONSENT
By signing this consent form, I confirm I have read the information in this consent form and have had the opportunity to ask questions. I will be given a signed copy of this consent form. I voluntarily agree to take part in this study.

____
Child's Name

____
Parent/Guardian's Name

_________________________ ____________
Parent/Guardian's Signature Date

____
Relationship to Child

____
Name of Person Obtaining Consent

_________________________ ____________
Signature of Person Obtaining Consent Date

APPENDIX C

ASSENT TO PARTICIPATE IN A RESEARCH STUDY

Assent to Participate in a Research Study

Why are we asking you to be in this research study?
We are asking you to take part in a research study because we are trying to develop a better equation that estimates your aerobic capacity (aerobic fitness) from tests used in physical education classes. This will help us better understand how fitness performance is related to overall health.

What happens in the research study?
If you agree to be in this study you will do the following:

- You will participate in a laboratory VO2 Peak test at the University of Utah.
- We will measure your body composition at the University of Utah by use of the BOD POD, skinfolds thickness, and BMI methods.
- You will participate in a one-mile run test and PACER tests on separate days.

Will any part of the research study hurt you?
No measures used in this study will hurt you. However, the VO2 Peak test will involve gradually increasing your running intensity on a treadmill until you reach exhaustion. During the test your heart rate will be monitored and pulmonary gas exchange (your breathing) will be measured by having you breath into a tube when you are running. You may feel emotionally uncomfortable as we measure your body fat % in lab settings. The BOD POD requires you to sit still in a closed chamber for 5 minutes and skinfolds thickness involves us gently pinching a fold of skin on your upper arm and lower leg to get a measurement. However, these tests will be administered in a relaxed and highly controlled environment **by a trained graduate student or faculty member as the same gender that you are (males will test males, females will test females)**. The one-mile run test and PACER test also involve running to exhaustion or until you voluntarily stop.

Will the research study help you or anyone else?
Being in this study will help us develop a new equation that predicts your aerobic capacity (aerobic fitness) for the FITNESSGRAM fitness assessment program. FITNESSGRAM assessments are used all across the world in physical education classes, therefore developing a new prediction equation has the potential to impact millions of youth on how they are classified (health risk, not a health risk) based on their performance on the fitness tests. You will also find out how you are classified by FITNESSGRAM by participating, which can be valuable by providing you information regarding your current health status.

Who will see the information about you?
All of your records about this research study will be kept locked up so no one else can see them except my faculty sponsor and myself. They will be kept in a password-protected computer that only the primary researcher can access.

What if you have any questions about the research study?
You can ask any questions that you have about the study. If you have a question later that you didn't think of now, you can call or e-mail me (Ryan D. Burns, 801-695-5693/Ryan.D.Burns@utah.edu) or ask me next time I see you.

Do you have to be in the research study?
If you don't want to be in this study, you don't have to be in it. Remember, being in this study is up to you and no one will be upset if you don't want to participate. You change your mind later if you want to stop. Please talk this over with your parents before you decide whether or not to participate. We will also ask your parents to give their permission for you to take part in this study. But even if your parents say "yes" you can still decide not to do this.

Agreeing to be in the study
Signing my name at the bottom (top two lines) means that I agree to be in this study. My parents and I will be given a copy of this form after I have signed it.

____
Printed Name

____________________________________ ____________________
Sign your name on this line | Date

____
Printed Name of Person Obtaining Assent

____________________________________ ____________________
Signature of Person Obtaining Assent | Date

The following should be completed by the study member conducting the assent process if the participant agrees to be in the study. Initial the appropriate selection:

_________ The participant is capable of reading the assent form and has signed above as documentation of assent to take part in this study.

_________ The participant is not capable of reading the assent form, but the information was verbally explained to him/her. The participant signed above as documentation of assent to take part in this study.

REFERENCES

American Alliance of Health, Physical Education, Recreation, and Dance. (1980). *Health related fitness test.* Reston, VA: Author.

Australian Council for Health, Physical Education and Recreation (ACHPER). (1996). *Handbook for the Australian Fitness Education Award manual.* Adelaide, SA: ACHPER Publications.

Biaggi, R. R., Vollman, M. W., Nies, M. A., Brener, C. E., Flakoll, P. J., Levenhagen, D. K., … Chen, K. Y. (1999). Comparison of air displacement plethysmography with dual-energy X-ray absorptiometry and 3 field methods for estimating body composition in middle-aged men. *American Journal of Clinical Nutrition*, *69*, 786−793.

Blair, S. N., Goodyear, N. N., Gibbons, L. W., & Cooper, K. H. (1984). Physical fitness and incidence of hypertension in healthy normotensive men and women. *Journal of the American Medical Association*, *252*, 487−490.

Blair, S. N., Kohl, H. W. III, Barlow, C. E., Paffenbarger, R. S., Gibbons, L. W., & Macera, C. A. (1995). Changes in physical fitness and all-cause mortality: A prospective study of healthy and unhealthy men. *Journal of the American Medical Association*, *273*, 1093−1098.

Blair, S. N., Kohl, H. W., Paffenbarger, R. S., Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical fitness and all-cause mortality: A prospective study of healthy men and women. *Journal of the American Medical Association*, *262*, 2395−2401.

Boiarskaia, E. A., Boscolo, M. S., Zhu, W., & Mahar, M. T. (2011). Cross-validation of an equating method linking aerobic FITNESSGRAM field tests. *American Journal of Preventive Medicine*, 41, S124−S130.

Boreham, C., Riddoch, C. (2001). The physical activity, fitness and health of children. *Journal of Sports Sciences*, *19*, 915−919.

Buono, M. J., Roby, F. G., Micale, F. G., Sallis, J. F., & Shepard, W. E. (1991). Validity and reliability of predicting maximum oxygen uptake via field tests in children and adolescents. *Pediatric Exercise Science*, *3*, 250−255.

Burke, E. M. (1976). Validity of selected laboratory and field tests of physical working capacity. *Research Quarterly*, *47*, 95−104.

Buskirk, E., & Taylor H. L. (1957). Maximal oxygen intake and its relation to body composition, with special reference to chronic physical activity and obesity. *Journal of Applied Physiology*, *11*, 72−78.

Caprio, S., Hyman, L. D., & Limb, C. (1995). Central adiposity and its metabolic correlates in obese adolescent girls. *American Journal of Physiology, 269*, E118−E126.

Carnethon, M. R., Gulati, M., & Greenland, P. (2005). Prevalence and cardiovascular disease correlates of low cardio-respiratory fitness in adolescents and adults. *Journal of the American Medical Association*, *294*, 2981−2988.

Castro-Piñero, J., Mora, J., Gonzalez-Montesinos, J. L., Sjöström, M., & Ruiz, J. R. (2009). Criterion-related validity of the one-mile run/walk test in children aged 8–17 years. *Journal of Sports Sciences*, *27*, 405−413.

Cataldo, R., John, J., Chandran, L., Pati, S., & Shroyer, A. L. W. (2013). Impact of physical activity intervention programs on self-efficacy in youths: A systemic review. *ISRN Obesity*, *2013*, 1-11. doi:10.1155/2013/586497.

Cale, L., Harris, J., & Chen, M. H. (2007). More than 10 years after "The Horse is Dead…": Surely it must be time to "Dismount"?! *Pediatric Exercise Science, 19,* 115−131.

Cole, T. J., Bellizi, M. C., Flegal, K. M., & Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: International survey. *British Medical Journal*, *320*, 1240−1243

Cooper, K. H. (1968). *Aerobics.* New York, NY: Bantam Books.

Cooper Institute for Aerobics Research (2004). *The Prudential FITNESSGRAM*: *Test administration manual* (3rd ed.). Champaign, IL: Human Kinetics.

Cureton, K. J., Baumgartner, T. A., & McManis, B. C. (1991). Adjustment of one mile walk test scores for skinfold thickness in youth. *Pediatric Exercise Science, 3,* 152–167.

Cureton, K., Bishop, P., & Huchinson, P., Newland, H., Vickery, S., & Zwiren, L. (1986). Sex differences in maximal oxygen uptake. Effect of equating haemoglobin concentration. *European Journal of Physiology and occupational Physiology*, *54*, 656−660.

Cureton, K. J., Boileau, R. A., Lohman, T. G., & Misner, J. E. (1977). Determinants of distance running performance in children: analysis of a path model. *Research Quarterly*, *48*, 270−279.

Cureton, K. J., Sloniger, M. A., O'Bannon, J. P., Black, D. M., & McCormick, W. P. (1995). A generalized equation for prediction of VO2 peak from 1-mile run/walk performance. *Medicine and Science in Sports and Exercise, 27*, 445−451.

Daniels, S. R. (2006). The consequences of childhood overweight and obesity. *Future Child*, *16*, 47−67.

Dewit, O., & Fuller, N. J. (2001). Whole-body air displacement plethysmography compared with hydrodensitomtery for body composition analysis. *Archives of Disease in Childhood*, *82*, 159−164.

Dwyer, T., & Blizzard, C. L. (1996). Defining obesity in children by biological endpoint rather than population distribution. *International Journal of Obesity*, *20*, 472−480.

Eisenmann, J. C. (2004). Physical activity and cardiovascular disease risk factors in children and adolescents: An overview. *Canadian Journal of Cardiology*, *20*, 295−301.

Eisenmann, J. C., Laurson, K. R, & Welk, G. J. (2011). Aerobic fitness percentiles for U.S. adolescents. *American Journal of Preventive Medicine, 41*, S106−S110.

Eisenmann, J. C., Welk, G. J., Ihmels, M., & Dollman, J. (2007). Fatness, fitness, and cardiovascular disease risk factors in children and adolescents. *Medicine and Science in Sport and Exercise*, *39*, 1251−1256.

Fields, D. A., Goran, M. I., & McCrory, M. A. (2002). Body-composition assessment via air-displacement plethysmography in adults and children: A review. *The American Journal of Clinical Nutrition*, *75*, 453−467.

Ford, E. (2005). Risks for all-cause mortality, cardiovascular disease, and diabetes associated with the metabolic syndrome: A summary of the evidence. *Diabetes Care*, *28*, 1769−1778.

Freedman, D. S., Khan, L. K., Dietz, W. H., Srinivasan, S. R, & Berenson, G. S. (2001). Relationship of childhood obesity and coronary heart disease risk factors in adulthood: The Bogalusa heart study. *Pediatrics*, *108*, 712−718.

Gahche, J., Fakhouri, T., Carroll, D. D., Burt, V. L., Wang, C. Y., & Fulton, J. E. (2014). Cardiorespiratory fitness levels among U.S. youth aged 12−15 years: United States, 1999−2004 and 2012. *National Center for Health Statistics Brief, 153*, 1−7.

Garn, S. M., Leonard, W. R., & Hawthorne, V. M. (1986). Three limitations of the body mass index. *American Journal of Clinical Nutrition*, *44*, 996−996.

Going, S., Lohman, T. G., Cussler, M. S., Williams, D. P., Morrison, J. A., & Horn, P. S. (2011). Percent body fat and chronic disease risk factors in U.S. children and youth. *American Journal of Preventative Medicine*, *41*, S77−S86.

Goran, M., Fields, D. A., Hunter, G. R., Herd, S. L., & Weinsier, R. L. (2000). Total body fat does not influence maximal aerobic capacity. *International Journal of Obesity*, *24*, 841−848.

Hannon, J. C., Ratliffe, T., & Williams, D. P. (2006). Agreement in body fat estimates between a hand-held bioelectrical impedance analyzer and skinfold thickness in African American and Caucasian adolescents. *Research Quarterly in Exercise and Sport*, *77*, 519−526.

Hartmann, D. P. (1977). Considerations in the choice of inter-observer reliability estimates. *Journal of Applied Behavioral Analysis*, 10, 103−116.

Himes, J. H. (2009). Challenges of accurately measuring and using BMI and other indicators of obesity in children. *Pediatrics*, *124*, S3−S22.

Ihmels, M., Welk, G. J., McClain, J. J., & Schaben, J. (2006). Reliability and convergent validity of field tests of body composition in young adolescents. *Journal of Physical Activity and Health*, *3,* S67−S77.

Iwaoka, H., Yokoyama, T., Nakayama, T., Matsumura, Y., Yoshitake, Y., Fuchi, T., …, & Tanaka, H. (1998). Determination of percent body fat by the newly developed sulfur hexafluoride dilution method and air displacement plethysmography. *Journal of Nutritional Science and Vitaminology (Tokyo)*, *44*, 561−568.

Jackson, A. S. (2006). The evolution and validity of health-related fitness. *Quest, 58*, 160−175.

Kearney, J. T., & Bynes, W. C. (1974). Relationship between running performance and predicted maximum oxygen uptake among divergent ability groups. *Research Quarterly*, *45*, 9−14.

Krahenbuhl, G. S., Pangrazi, R. P., Burkett, L. N., Schneider, M. J., & Petersen, G. (1977). Field estimation of VO2max in children eight years of age. *Medicine and Science in Sports*, *9*, 37–40.

Kraus, H., & Hirschland R. P. (1953). Muscular fitness and health. *Journal of Physical Education Recreation and Dance*, *24,* 17−19.

Kraus, H., & Hirschland, R. P. (1954), Minimum muscular fitness tests in school children. *Research Quarterly*, *25*, 178−188.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159−174.

Laurson, K. R., Eisenmann, J. C., & Welk, G. J. (2011a). Development of youth percent body fat standards using receiver operating characteristic curves. *American Journal of Preventive Medicine, 41*, S93−S99.

Laurson, K. R., Eisenmann, J. C., & Welk, G. J. (2011b). Body fat percentile curves for U.S. children. *American Journal of Preventive Medicine, 41*, S87−S92.

Laurson, K. R., Eisenmann, J. C., & Welk, G. J. (2011c). Body Mass Index standards based on agreement with health-related body fat. *American Journal of Preventive Medicine, 41*, S100−S105.

Lee, C. D., Blair, S. N., & Jackson, A. S. (1999). Cardio-respiratory fitness, body composition, and all-cause and cardiovascular disease mortality in men. *American Journal of Clinical Nutrition, 69*, 373–80.

Leger, L., & Lambert, J. (1982). A maximal multistage 20-m shuttle run test to predict VO_{2MAX}. *European Journal of Applied Physiology*, *49*, 1−12.

Leger, L. A., Mercier, D., Gadourey, C., & Lambert, J. (1988). The multistage 20-metre shuttle run test for aerobic fitness. *Journal of Sport Science*, *6*, 93−101.

Lintsi, M., Kaarma, H., & Kull, I. (2004). Comparison of hand-to-hand bio-imp edance and anthropometry equations versus dual-energy X-ray absorptiometry for the assessment of body fat percentage in 17-18-y ear old conscripts. *Clinical Physiology and Functional Imaging*, *24*, 85−90.

Liu, N. Y-S., Plowman, S. A., & Looney, M. A. (1992). The reliability and validity of the 20-meter shuttle test of aerobic fitness in adolescent schoolchildren. *Journal of Sports Medicine and Physical Fitness*, *30*, 19−23.

Lobelo, F., & Ruiz, J. R. (2007). Cardiorespiratory fitness as criterion validity for health-based metabolic syndrome definition in adolescents. *Journal of the American College of Cardiology*, *50*, 471−472.

Lukaski, H. C., & Siders, W. A. (2003). Validity and accuracy of regional bioelectrical impedance devices to determine whole-body fatness. *Nutrition*, *19*, 1851–1857, 2003.

Maffeis C., Schutz Y., Schena, F., Zaffanello, M., & Pinelli, L. (1993) Energy expenditure during walking and running in obese and nonobese prepubertal children. *Journal of Pediatrics*, *123*, 193−199.

Mahar, M. T., Guerieri, A. M., Hanna, M. S., & Kemble, D. (2011). Estimation of aerobic fitness from 20-m multistage shuttle run test performance. *American Journal of Preventative Medicine*, *41*, S117−S123.

Mahar, M. T., Welk, G. J., Rowe, D. A., Crotts, D. J., & McIver, K. L. (2006). Development and validation of a regression model to estimate VO_{2PEAK} from PACER 20-m shuttle run performance. *Journal of Physical Activity and Health*, *3,* S34−S46.

Mahoney, C. (1992). 20-MST and PWC_{170} validity in non-Caucasian children in the UK. *British Journal of Sports Medicine*, *26*, 45−47.

Malina, R. M. (1996). Tracking of physical activity and physical fitness across the lifespan. *Research Quarterly for Exercise and Sport*, *67*, S48−S57.

Malina, R. M., Bouchard, C., & Bar-Or, O. (2004). *Growth, maturation, and physical activity* (2nd ed.). Champaign, IL: Human Kinetics.

Matsuzaka, A., Takahashi, Y., Yamazoe, M. Kumakura, N., Ikeda, A., & Wilk, B. (2004). Validity of the multistage 20-m shuttle run test for Japanese children, adolescents, and adults. *Pediatric Exercise Science*, *16*, 113−125.

McArdle, W. D., Katch, F. L., & V. L. Katch (1999). *Essentials of Epidemiology* (3rd ed.). Baltimore, MD: Lippincott Williams & Wilkens.

McCrory, M. A., Gomez, T. D., Bernauer, E. M., & Mole, P. A. (1995). Evaluation of a new air displacment plethysmograph for measuring human body composition. *Medicine and Science in Sports and Exercise*, *27*, 1686−1691.

Meredith, M., & Welk, G. J. (2010). *Fitnessgram/Actvitygram test administration manual* (Updated 4th ed). Champaign, IL: Human Kinetics.

Moreno, L. A., Mesana, M. I., Gonzalez-Gross, M., Gil, C. M., Fleta, J., Warnberg, J., … Bueno, M. (2006). Anthropometric body fat composition reference values in Spanish adolescents. The AVENA study. *European Journal of Clinical Nutrition*, *60*, 191−196.

Morrison, J. A., Barton, B. A., Biro, F. M., Daniels, S. R., & Sprecher, D. L. (1999). Overweight, fat patterning, and cardiovascular disease risk factors in black and white boys. *Journal of Pediatrics*, *135*, 451−457.

Morrow, J. R., Zhu, W., Franks, B. D., Meredith, M. D., & Spain, C. (2009). 1958-2008: 50 years of youth fitness tests in the U.S. *Research Quarterly in Exercise and Sport*, *80*, 1−11.

Must, A., Jacques, P. F., Dallal, G. E., Bajema, C. J., & Dietz, W. H. (1992). Long-term morbidity and mortality of overweight adolescents. A follow-up of the Harvard Growth Study of 1922 to 1935. *North England Journal of Medicine*, *327*, 1350−1355.

National Association for Sport and Physical Education (2011). *Physical Education for Lifelong Fitness* (3rd Ed). Champaign, IL: Human Kinetics.

Nunez, C., Kovera, A. J., Pietrobelli, A., Heshka, S., Horlick, M., Kehayias, J. J., … Heymsfield, S. B. (1999). Body composition in children and adults by air displacement plethysmography. *European Journal of Clinical Nutrition*, *53*, 382−387.

O'Connor, D. P., Mahar, M. T., Laughlin, M. S., & Jackson, A. S. (2011). The Bland-Altman method should not be used in regression cross-validation studies. *Research Quarterly for Exercise and Sport*, *82*, 610−616.

Ogden, C. L., Carroll, M. D., Curtin, L. R., Lamb, M. M., & Flegal, K. M. (2010). *Journal of the American Medical Association*, *303*, 242-249.

Ogden, C. L., Kuczmarski, R. J., Flegal, K. M., mei, Z., Guo, S., Wei, R., …, &Johnson, C. L. (2002). Centers for Disease Control and Prevention 2000 growth charts for the United States: improvements to the 1977 National Center for Health Statistics version. *Pediatrics*, *109*, 45−60.

Oppliger, R. A., Clark, R. R., & Kuta, J. M. (1992). Efficacy of skinfold training clinics: A comparison between clinic trained and experienced testers. *Research Quarterly for Exercise and Sport, 63,* 438−443.

Ortega, F. B., Ruiz, J. R., Castillo, M. J., & Sjostrom, M. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity, 32*, 1−11.

Pagano, M., & Gauvreau, K. (2000). *Principles of biostatistics* (2nd ed.). Belmont, CA: Brooks/Cole, Cengage Learning.

Pate, R. R., Wang, C. Y., Dowd, M., Farrell, S. W., & O'Neill, J. R. (2006). Cardiorespiratory fitness levels among U.S. youth 12 to 19 years of age: Findings from the 1999−2002 National Health and Nutrition Examination Survey. *Archives of Pediatric and Adolescent Medicine*, *160*, 1005−1012.

Prentice, A. M., & Jebb, S. A. (2001). Beyond body mass index. *Obesity Review*, *2*, 141-147.

Safrit, M. J. (1990). The validity and reliability of fitness tests for children: A review. *Pediatric Exercise Science*, *2*, 9−28.

Safrit, M. J., Baumgartner, T. A., Jackson, A. S, & Stamm, C. L. (1980). Issues in setting motor performance standards. *Quest*, *32*, 152−162.

Savva, S. C., Tornaritis, M., & Savva, M. E. (2000). Waist circumference and waist-to-height ratio are better predictors of cardiovascular disease risk factors in children than body mass index. *International Journal of Obesity and Related Metabolic Disorders*, *24*, 1453–1458.

Schwimmer, J. B., Burwinkle, T. M., & Varni, J. W. (2003). Health-related quality of life for severely obese children and adolescents. *Journal of the American Medical Association*, *289*, 1813−1819.

Shaw, V. W. (1986). The accuracy of two training methods on skinfold assessment. *Research Quarterly for Exercise and Sport, 57*, 85−90

Slaughter, M. H., Lohman, T. G., & Boileau, R. A. (1988). Skinfold equations for estimation of body fatness in children and youth. *Human Biology*, *60*, 709−723.

Sui, X., Hooker, S. P., & Lee, I. M. (2007). A prospective study of cardio-respiratory fitness and and risk of type 2 diabetes in women. *Diabetes Care, 31*, 550–555.

Thakur, J. S., Yadav, R. C., & Singh, V. K. (2010). Influence of body composition on the dimensions of VO2 PEAK. *VSRD Technical & Non-Technical Journal*, *1*, 72−77.

The President's Council on Physical Fitness and Sports (2007). The President's Challenge: The Health Fitness Test. Retrieved from http://www.presidentschallenge.org/ educators/program_details/health_fitness_test.aspx.

Trowbridge, C. S., Gower, B. A., Nagy, T. R., Hunter, G. R., Treuth, M. S., & Goran, M. I. (1997). Maximal aerobic capacity in African-American and Caucasian prepubertal children. *American Journal of Physiology Endocrinology and Metabolism*, *273*, 809−814.

Vincent, S. D., Barker, R., Clarke, M., & Harrison, J. (1999). A comparison of peak heart rates elicited by the 1-mile run/walk and the Progressive Aerobic Capacity Endurance Run. *Research Quarterly in Exercise and Sport*, *70*, 75−78.

Wang, G., & Dietz, W. H. (2002). Economic burden of obesity in youths aged 6 to 17 years: 1979−1999. *Pediatrics*, 109(5), E81.

Washino, K., Takada, H., Nagashima, M., & Iwata, H. (1999). Significance of the atherosclerogenic index and body fat in children as markers for future, potential coronary heart disease. *Pediatric International*, *41*, 260−265.

Welk, G. J., Going, S. B., Morrow, J. R., Jr., & Meredith, M. D. (2011). Development of new criterion-referenced fitness standards in the FITNESSGRAM program: Rationale and conceptual overview. *American Journal of Preventive Medicine*, *41*, S63−S67.

Welk, G. J., Laurson, K. R., Eisenmann, J. C., & Cureton, K. J. (2011). Development of youth aerobic-capacity standards using receiver operating characteristic curves. *American Journal of Preventive Medicine, 41*, S111−S116.

Welk, G. J., Saint-Maurice Maduro, P. F., Laurson, K. R., & Brown, D. D. (2011). Field evaluation of the new FITNESSGRAM criterion-referenced standards. *American Journal of Preventative Medicine*, *41*, S131−S142.

Wells, J. C., & Fuller, N. J. (2001). Precision of measurement and body size in whole-body air-displacement plethysmography. *International Journal of Obesity and Related Metabolic Disorders*, *25*, 1161−1171.

Welsman, J. R., Bywater, K., Farr, C., Welford, D., & Armstrong, N. (2005). Reliability of peak VO_2 and cardiac output assessed using thoracic impedance in children. *European Journal of Applied Physiology*, *94,* 228−234.

Williams, D. P., Going, S. B., Lohman, T. G., Harsha, D. W., Srinivasan, S. R., Webber, L. S., & Berenson, G. S. (1992). Body fatness and risk for elevated blood pressure, total cholesterol, and serum lipoproteins ratios in children and adolescents. *American Journal of Public Health*, *82*, 358−363.

Williams, J., Wake, M., Hesketh, K., Mahar, E., & Waters, E. (2005). Health-related quality of life for overweight and obese children. *Journal of the American Medical Association*, *293*, 70−76.

Zhu, W., Mahar, M. T., Welk, G. J., Going, S. B., & Cureton, K. J. (2011). Approaches for development of criterion-referenced standards in health-related youth fitness tests. *American Journal of Preventive Medicine*, *41,* S68−S76.

Zhu, W., Plowman, S. A., & Park, Y. (2010). A primer-test centered equating method for setting cut-off scores. *Research Quarterly for Exercise and Sport*, *81*, 400−409.

Zhu, S., Wang, Z., Shen, W., Heymsfield, S. B., & Heshka, S. (2003). Percentage body fat ranges associated with metabolic syndrome risk: Result based on the third National Health and Nutrition Examination Survey (1988−1994). *American Journal of clinical Nutrition*, *78*, 228−235.