Measurement agreement of FITNESSGRAM aerobic capacity and body composition standards

by

Katelin M. Blasingame

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Kinesiology

Program of Study Committee:
Gregory J. Welk, Major Professor
Spyridoula Vazou
Michelle Ihmels

Iowa State University
Ames, Iowa

2012

Copyright © Katelin M. Blasingame, 2012. All rights reserved.
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................ iv
LIST OF TABLES ........................................................................................................... v
ACKNOWLEDGMENTS ................................................................................................. vi
ABSTRACT ......................................................................................................................... vii
CHAPTER 1: INTRODUCTION ....................................................................................... 1
  Specific Research Questions ....................................................................................... 4
CHAPTER II: EXTENDED LITERATURE REVIEW ........................................................ 6
  Physical Activity, Physical Fitness and Health ......................................................... 8
  Overview of the FITNESSGRAM Test Battery ....................................................... 9
    Aerobic Capacity .................................................................................................. 10
    Body Composition ............................................................................................... 12
  FITNESSGRAM Criterion-Referenced Standards .................................................. 16
    Development of the New Body Composition Standards .................................. 17
    Development of the New Aerobic Capacity Standards ................................... 19
  Evaluating Measurement Agreement ..................................................................... 20
CHAPTER III: METHODS ............................................................................................... 22
  Participating Schools ............................................................................................... 22
  Instruments ............................................................................................................... 22
    Aerobic Capacity Assessments ......................................................................... 22
    Body Composition Assessments ....................................................................... 23
  Data Collection Procedures ..................................................................................... 24
  Statistical Analysis ................................................................................................. 26
CHAPTER IV: RESULTS ................................................................................................. 28
  Aerobic Capacity ..................................................................................................... 28
  Descriptive Statistics ............................................................................................. 28
  Evaluation of Measurement Agreement ............................................................... 28
  Evaluation of Classification Agreement ............................................................... 29
  Body Composition ................................................................................................. 30
Descriptive Statistics ........................................................................................................ 30
Evaluation of Measurement Agreement ........................................................................ 30
Evaluation of Classification Agreement ........................................................................ 31

CHAPTER V: DISCUSSION ........................................................................................................... 33
Discussion of Aerobic Capacity Findings ....................................................................... 33
Discussion of Body Composition Findings ..................................................................... 36
Summary of Results and Implications for Youth Fitness Testing .................................. 40
Conclusion ......................................................................................................................... 43

REFERENCES ...................................................................................................................... 45

TABLES AND FIGURES ......................................................................................................... 52

APPENDIX ............................................................................................................................... 70
Panel A: Testing Procedures for the PACER ................................................................. 70
Panel B: Testing Procedures for the One-Mile Run ........................................................ 73
Panel C: Testing Procedures for Skinfold Measurements ................................................ 76
Panel D: Testing Procedures for Portable Bioelectric Impedance Analyzers ................ 77
Panel E: Testing Procedures for Body Composition ....................................................... 78
LIST OF FIGURES

Figure 1. Males BMI Standards .................................................................52
Figure 2. Females BMI Standards ............................................................53
Figure 3. Males Body Fat Standards ........................................................54
Figure 4. Females Body Fat Standards ....................................................55
Figure 5. Males Aerobic Capacity Standards ...........................................56
Figure 6. Females Aerobic Capacity Standards .......................................57
Figure 7. Mean of the Difference Between Measures of Aerobic Capacity ....58
Figure 8. Bland-Altman Plot for Aerobic Capacity ...................................59
Figure 9. Mean of the Difference Between Measures of Percent of Body Fat ....60
Figure 10. Bland-Altman Plot for Percent of Body Fat ..............................61
LIST OF TABLES

Table 1. Aerobic Capacity Descriptives ..........................................................62
Table 2. Aerobic Capacity Zone Classification ...............................................63
Table 3. Aerobic Capacity Matrix of Classification Agreement .....................64
Table 4. Aerobic Capacity Kappa Scores .......................................................65
Table 5. Body Composition Descriptives .......................................................66
Table 6. Body Composition Zone Classification ..........................................67
Table 7. Body Composition Matrix of Classification Agreement ..................68
Table 8. Body Composition Kappa Scores .....................................................69
ACKNOWLEDGMENTS

It is my pleasure to thank those who made this thesis possible,
my POS committee members, especially my major professor Dr. Greg Welk,
the research team involved in this project, who dedicated much time and expertise,
my friends, family and Tyler, who encouraged and supported me every step of the way.
Thank you
ABSTRACT

Introduction: FITNESSGRAM is a fitness testing program developed by The Cooper Institute. The FITNESSGRAM program provides two different tests for assessing aerobic capacity (VO₂ max), the PACER and the one-mile run. The FITNESSGRAM program also provides two major options for assessing body composition, body fat using skinfold measurements or bioelectrical impedance analyzers (BIA) and body mass index (BMI). Criterion-referenced standards are used in FITNESSGRAM to determine the child’s level of fitness compared to the level of fitness needed for good health. New standards were recently developed that classify student’s fitness level into 3 zones based on potential risk for developing metabolic syndrome (Healthy Fitness Zone (low risk), Needs Improvement – Low Risk (moderate risk), and Needs Improvement – High Risk (high risk)). The new standards were also designed to improve classification agreement between alternative fitness assessments. Preliminary results have supported the utility of the new standards but additional validation research is needed. Purpose: The present study examined the measurement agreement and classification agreement between alternative indicators of aerobic capacity and body composition assessments in FITNESSGRAM. Methods: Data were collected in partnership with a local school district that agreed to let researchers administer aerobic capacity and body composition tests as part of normal physical education testing. Data were collected at a high school (grades 9 – 12) and a junior high school (grades 7, 8). Data were combined across grades to enable measurement agreement to be examined at both the junior high school and high school level. Separate two-way (gender x school type) ANOVA analyses were used to evaluate grade and gender differences in fitness levels with
both sets of indicators (4 total ANOVA analyses). Percent agreement was computed between the assessments and kappa statistics were used to understand the practical differences in agreement. **Results:** A total of 701 students had data for both measures of aerobic capacity (PACER and one-mile run) and 889 had all measures of body composition (BMI and body fat). The correlation between predicted aerobic capacity from the PACER and one-mile run tests was high \( r = 0.94 \). The percent of students achieving the HFZ was similar for the one-mile run (69.6%) and the PACER (72.6%). The total percent classified similarly by the two tests was 92.7%. The kappa score was also very good (0.84). The overall correlation between predicted measure of percent of body fat using the BIA and skinfold was high (0.86). The percent of students achieving the HFZ was BIA (58.8%), skinfold (61.3%), and BMI (55.2%). The classification agreement averaged 70.6% for the various pairwise comparisons (skinfold/BIA: 76.2%, skinfold/BMI: 68%, BIA/BMI: 67.7%). The classification agreement based on kappa scores were fair to moderate (skinfold/BIA: 0.575, skinfold/BMI: 0.454, BIA/BMI: 0.462). **Conclusion:** The findings of this study suggest that there is strong agreement between measures of aerobic capacity, while there are moderate levels of agreement between measures of body composition.
CHAPTER 1: INTRODUCTION

In the past decade there has been a significant increase in the prevalence of overweight children and adolescents, with no evidence of any reversal in these trends (Ogden, Carroll, Curtin, McDowell, Tabak and Flegal, 2006; Centers for Disease Control and Prevention, 2010; Ogden, Carroll, Curtin, Lamb and Flegal, 2010). It has been found that those who are overweight as youth are more likely to remain overweight into adulthood (Freedman, Khan, Serdula, Dietz, Srinivasan and Berenson, 2005). Given that youth spend such a large portion of their day in school, schools are often looked to as a way to help implement and support a solution to assist in reversing these trends of being overweight. Since fitness testing is a common component of most physical education programs, increased attention is being placed on the results of school based fitness testing (Lee, Burgeson, Fulton and Spain, 2007).

FITNESSGRAM is a fitness testing program developed by The Cooper Institute in 1982. The program focuses on health-related fitness and includes protocols for the assessment of aerobic capacity, muscular strength, muscular endurance, flexibility and body composition (Meredith and Welk, 2010). While each dimension is important for health, most public health attention has focused on aerobic capacity and body composition. Aerobic capacity and body composition are both especially important because they are linked to risk for various chronic diseases including coronary heart disease, hypertension, and type 2 diabetes mellitus (Anderssen, Cooper, Riddoch, Sardinha, Harro, Brage and Andersen, 2007; US Department of Health and Human Services, 2008).
The FITNESSGRAM program provides teachers with several options for assessing aerobic fitness and body composition. The FITNESSGRAM program provides two different tests for assessing aerobic capacity (VO₂ max). The 1) PACER (Progressive Aerobic Cardiovascular Endurance Run) test is a progressive 20-meter shuttle run test that challenges students to run at progressively faster paces until they can’t keep up. The second test is the 2) one-mile run where students run one-mile as fast as possible. The PACER is the recommended test; however the one-mile run is also widely used. The FITNESSGRAM program provides two major options for assessing body composition (1) body fat and 2) body mass index. Skinfold measurements and bioelectrical impedance analyzers (BIA) are used to measure body fat percentages. Body mass index (BMI) calculates an individual’s weight relative to height.

Both assessments are evaluated using criterion-referenced standards that reflect the amount of fitness needed for good health. Criterion-referenced standards are more useful for fitness evaluation since it makes it possible for individuals to compare their overall fitness to an absolute criterion (The Cooper Institute, 2011). These standards help to place the individual in either the Health Fitness Zone (HFZ) or the Needs Improvement Zone (NI). The HFZ represents those individuals who are considered to be at a low risk for developing metabolic syndrome, while those who are classified as NI are considered to be at a higher risk for developing metabolic syndrome.

Research has been conducted on the reliability and validity of the various aerobic fitness tests e.g. (Leger, Mercier, Gadoury and Lambert, 1988; Liu, Plowman and Looney, 1992; George, Vehrs, Allsen, Fellingham and Fisher, 1993; Plowman and Liu, 1999; Mahar, Guerieri, Hanna and Kemble, 2011) as well as the body composition tests e.g. (Lukaski,
Bolonchuk, Hall and Siders, 1986; Jackson, Pollock, Graves and Mahar, 1988; Heitmann, 1990; Gutin, Litaker, Islam, Manos, Smith and Treiber, 1996; Pietrobelli, Faith, Allison, Gallagher, Chiumello and Heymsfield, 1998; Jebb, Cole, Doman, Murgatroyd and Prentice, 2000; Cable, Nieman, Austin, Hogen and Utter, 2001; Utter, Scott, Oppliger, Visich, Goss, Marks, Nieman and Smith, 2001; Yeung and Hui, 2010). The health related standards have been well supported in the literature. For example, Pate et al. (2006) demonstrated that the aerobic capacity standards discriminated between youth that were at risk versus those not at risk for chronic disease. This discrepancy is a main goal of health-related fitness testing.

While the standards have had good utility, some inconsistencies have been noted. For example, data from a large statewide evaluation in Texas (Welk, Meredith, Ihmels and Seeger, 2010) revealed differences in fitness achievement depending on what test was used. Gender differences in aerobic fitness tests were also evident. While there were equal numbers of males and females reaching the HFZ in the one-mile run, there were far more females than males reaching the HFZ in the PACER test. Inconsistencies have also been noted with regard to the body composition assessments. Less than 10% of youth were classified in the NI zone, with it being easier for females to reach the HFZ than males (Welk, Going, Morrow and Meredith, 2011). Although it is entirely possible for less than 10% of our youth to be in the NI, this is not represented in current research.

New standards were recently developed to address these concerns. The new standards used nationally representative data from the National Health and Nutrition Examination Survey (NHANES) and used more advanced statistical methods. A key difference with the standards is that there are now three fitness zones rather than two. Previous FITNESSGRAM standards used a single threshold to divide students into either the HFZ or the NI. The new
standards divide the NI into two distinct zones, NI – high risk or NI – low risk (The Cooper Institute, 2011). This provides more effective feedback since the messages can reflect the chance of developing certain health implications such as the risk of developing metabolic syndrome for each zone, as well as tips on how to improve the child’s overall health.

The new standards provide many advantages over the previous standards but additional research is needed to test out their utility in practice. One key component is to test whether there is good classification agreement between the alternative assessments of aerobic fitness and body composition. This was a high priority for establishing the new standards making it important to evaluate the agreement in independent samples of youth. The present study will compile data from a large sample of youth to determine the percent that achieve the criterion-referenced standards using the various assessments. Results will be examined by age and gender to provide a comprehensive reporting of the relationships.

**Specific Research Questions**

The study will evaluate two specific research questions related to the new FITNESSGRAM standards.

1. A. Do youth receive similar results if they are assessed with the new standards for the PACER test and the one-mile run test?

   B. Are there gender and/or age differences associated with the classification agreement between the PACER test and the one-mile run test?

   It was hypothesized that there would be good overall agreement between the PACER test and one-mile run test classifications. It is further hypothesized that agreement will be similar for males and females of different ages.
2. A. Do youth receive similar results if they are assessed with the new standards for skinfold measurements, BIA and BMI?

B. Are there gender and/or age differences associated with the classification agreement between skinfold measurements, BIA and BMI?

It was hypothesized that there would be good overall agreement between skinfold measurements, BIA and BMI classifications. It is further hypothesized that agreement will be similar for males and females of different age.
CHAPTER II: EXTENDED LITERATURE REVIEW

Fitness testing is a common component of most physical education (PE) programs across the nation (Lee, Burgeson et al., 2007). Fitness testing provides teachers with an easy and effective way to assess a number of health related variables over a short period (Freedson, Cureton and Heath, 2000). Fitness testing is used in schools for a variety of reasons. However, according to the National Association for Sports and Physical Education (NASPE), “The primary goal of assessment should be seen as the enhancement of learning, rather than simply the documenting of learning” (NASPE, 1995). There has been considerable debate about whether school-based fitness testing is a good use of time in the curriculum (Mahar and Rowe, 2008; Silverman, Keating and Phillips, 2008; Welk, 2008; Morrow and Ede, 2009). Some of the main uses of fitness testing include facilitating in fitness education, providing feedback and allowing personal tracking (Welk and Meredith, 2008).

While fitness testing continues to be the mainstay of most PE programs, some have questioned the utility and purpose of standardized testing (Corbin, Pangrazi and Welk, 1995; Rowland, 1995; Cale, Harris, Chen, Corbin, Fox, Morrow and Plowman, 2007). A recent supplement of “Measurement in Physical Education and Exercise Science” characterized different perspectives about youth fitness testing. Mahar and Rowe (2008) described a three-stage paradigm representing the validation process used in fitness-related research. They emphasized the importance of reliable and valid fitness testing techniques and guidelines, especially the meaningfulness of confidence in the testing process and why it is needed to ensure accurate data. Silverman, Keating, and Phillips (2008) addressed the association
attributed with fitness testing, and how it has the potential to be a positive experience for everyone involved, especially youth. They also incorporated ways to use fitness testing as a method to enhance the instruction of fitness in a physical education setting, to help increase youth’s positive attitude, knowledge and understanding of physical activity.

Wiersma and Sherman (2008) focused on the psychological aspects associated with youth and their fitness testing performance. They pointed out that it is extremely important to minimize the negative connotations and experiences that can be associated with improper use and implementation of fitness testing, and focus on the positive experiences such as maximizing effort, enjoyment, and motivation. Welk (2008) discussed the important role that physical activity assessments play in promoting involvement in physical activity. He also described the importance and relevance of knowledge and development of behavioral skills that will help youth establish lifelong patterns of physical activity in order to help maintain a healthy and active lifestyle into their later years. Collectively, the results of this supplement emphasize that youth fitness testing can provide valuable information if used properly within a quality physical education program.

This literature review will summarize issues associated with fitness testing and provide a rationale for the proposed research. The first section will review the relationships between physical activity, physical fitness, and health and explain why school based fitness testing is so important. The second section will summarize the elements of the FITNESSGRAM youth fitness battery to be evaluated in the present study. Particular attention will be paid to the options available to assess aerobic capacity and body composition given that they are the focus of this project. The third section will explain the process used in FITNESSGRAM to establish criterion-referenced standards. This section will
provide a background on the new standards as well as comparisons with the previous standards. The final section will evaluate some principles of test agreement, and review previous studies that have examined agreement between measures of fitness in youth.

**Physical Activity, Physical Fitness and Health**

Physical activity and physical fitness are often used interchangeably, but they refer to two distinct concepts. Physical activity is defined as a bodily movement that results in energy expenditure, while physical fitness is characterized as a set of skills- or health-related qualities (Caspersen, Powell and Christenson, 1985). The World Health Organization (1948) defines health as a complete physical, mental, and social well-being and not merely just the absence of sickness or disease. Participation in physical activity is known to lead to improvements in physical fitness but associations tend to be weaker in youth than in adolescents or adults (Morrow, 2005). Many studies have reported significant associations, but some studies have reported that genetics and maturation have a strong influence on fitness during childhood and early adolescence (Ekelund, Poortvliet, Nilsson, Yngve, Holmberg and Sjostrom, 2001; Dencker, Thorsson, Karlsson, Linden, Svensson, Wollmer and Andersen, 2006; Dencker, Thorsson, Karlsson, Linden, Wollmer and Andersen, 2008; Kristensen, Moeller, Korsholm, Kolle, Wedderkopp, Froberg and Andersen, 2010). Both physical activity and fitness have important influences on health. Physical activity and fitness are typically found to have independent effects on a variety of health-related occurrences, such as being overweight, obese, or having coronary heart disease (Katzmarzyk, Malina and Bouchard, 1999; Kim, Must, Fitzmaurice, Gillman, Chomitz, Kramer, McGowan and Peterson, 2005; Mojica, Poveda, Pinilla and Lobelo, 2008)
Within the past decade, there has been a significant increase in the prevalence of overweight youth and adolescents, and there is no evidence of a reversal of these trends (Ogden, Carroll et al., 2006; Centers for Disease Control and Prevention, 2010; Ogden, Carroll et al., 2010). The increasing epidemic of obesity in youth is of considerable public health concern since youth that are overweight are more likely to remain overweight into adulthood (Freedman, Khan et al., 2005). Schools have not been implicated as a cause of the obesity epidemic, but they are often viewed as an important part of the solution. The increased attention on fitness in youth has brought greater attention on the results of school-based fitness tests.

**Overview of the FITNESSGRAM Test Battery**

FITNESSGRAM is a fitness testing program developed by The Cooper Institute in 1982. The FITNESSGRAM testing battery includes a variety of reliable and valid field assessments for the primary dimensions of health-related physical fitness (aerobic capacity, muscular strength, muscular endurance, flexibility and body composition). The associated software and web-based tool enables data to be entered by students, or by teachers. The program also produces individualized feedback on ways to improve or enhance an individual’s overall fitness. Another beneficial aspect of FITNESSGRAM is that it allows teachers to print off easy to read personalized reports that briefly explain each individual test and how the student did in comparison with the standards. The report also includes feedback on ways to help improve the students overall fitness (The Cooper Institute, 2011). The present section will describe the nature of the components that make up FITNESSGRAM.

Teachers have options for assessing the different dimensions of fitness. For aerobic capacity teachers can choose the PACER (Progressive Aerobic Cardiovascular Endurance
Run) test, the one-mile run, or the walk test. For body composition, teachers can choose to report estimates of body fatness (typically obtained with skinfold calipers or bioelectric impedance analyzer) or compute body mass index (BMI) using height and weight data. Muscular strength and endurance measures include the curl-up, trunk lift, push-up, pull-up, and flexed arm hang. The final component is flexibility, which can be measured using the back savor sit and reach, or the shoulder stretch (Meredith and Welk, 2010). The study will focus on aerobic capacity and body composition so additional detail is provided on these assessment tools.

**Aerobic Capacity**

Aerobic capacity represents the amount of oxygen the body can take in, transport, and then utilize during exercise. Cardiovascular fitness is important because it is linked to a reduction in fatigue, lower risk of coronary heart disease, hypertension, type 2 diabetes mellitus, and a variety of other chronic diseases (Anderssen, Cooper et al., 2007; US Department of Health and Human Services, 2008). There has been a vast amount of research showing the inverse relationship between high cardiovascular fitness and a decrease in cardiovascular morbidity during childhood and adulthood (Ferreira, Twisk, Stehouwer, Van Mechelen and Kemper, 2003; Zeno, Kim-Dorner, Deuster, Davis, Remaley and Poth, 2010). Although FITNESSGRAM offers three different types of tests to determine youth’s level of aerobic capacity, for the purpose of this study we focused on two, the PACER test and the one-mile run.

The PACER test was adapted from the 20-meter shuttle run (Leger, Mercier et al., 1988). The test is set at a time cadence, in which it begins at an easy steady pace and gradually get more difficult each minute as time progresses. The PACER test is the
recommended test in the FITNESSGRAM battery since it has a number of important advantages. It can be conducted indoors with limited space and does not require self-monitoring (Freedson, Cureton et al., 2000). Other advantages of the PACER test include a built-in warm-up, the fact that it helps students learn how to pace themselves, and a key psychological advantage: students who have poorer performance finish first as opposed to last with other tests (Meredith and Welk, 2010).

The one-mile run test also measures aerobic capacity and is an alternative to the PACER test. The objective of this test is to run one-mile as fast as possible. If a student is not able to run the entire distance, walking is permitted. In order to conduct the test it is important to have a flat running course as well as a way for students to easily count laps. It is recommended to inform students about the importance of pacing, as well as provide students with an opportunity to practice prior to the test. If a student does not properly pace themselves, they may begin running too quickly at the beginning of the mile and then be forced into a walk towards the end. In order to get the most accurate calculation for aerobic capacity, the student needs to run at a steady pace during the entire test. The student’s aerobic capacity is determined using an established prediction equation that takes into account age, gender, body mass index and one-mile run time (Meredith and Welk, 2010).

Past studies have examined the reliability and validity of the aerobic fitness tests. In order to assure FITNESSGRAM tests were valid, studies compared predicted VO$_2$ calculated using FITNESSGRAM tests to observed VO$_2$ assessed on a treadmill. Liu, Plowman, and Looney (1992) conducted a study testing the 20-meter shuttle test, which is the basis for the PACER test, and found the test to be reliable as well as valid. They evaluated the test-retest reliability, the concurrent validity in number of laps, as well as the prediction equation for
VO$_2$ max created by Leger in a group of 12-15 year old children. An intra-class coefficient of 0.93 was calculated on 20 students. VO$_2$ peak was then obtained by a treadmill test on 48 subjects. VO$_2$ peak was found to be significantly correlated with the number of laps completed on the treadmill in males (n = 22; r = 0.65), females (n = 26; r = 0.51), and males and females (r = 0.69).

The accuracy of the one-mile run test to determine VO$_2$ max has also been tested. It was determined that the one-mile run test can accurately predict aerobic capacity (George, Vehrs et al., 1993). George et al (1993) measured VO$_2$ max using the treadmill protocol and then compared it to the submaximal one-mile run estimation of VO$_2$ max. A cross validation compared observed and predicted relative VO$_2$ max and found there to be a $r_{adj} = 0.84$ and standard error of estimate = 3.1 ml kg$^{-1}$ min$^{-1}$. With $p > 0.05$ the observed VO$_2$ max 47.4 ± 5.9, and predicted VO$_2$ max 47.5 ± 5.1, were not found to be statistically significant. This suggests that the one-mile run test provides an accurate field assessment of aerobic capacity.

**Body Composition**

Body composition provides the individual with an estimated amount of their overall percent of body fat. Body composition is important because it is also linked to risk factors of coronary heart disease, hypertension, type 2 diabetes mellitus, and a variety of other chronic diseases (US Department of Health and Human Services, 2008). Although body composition needs to be carefully administrated for privacy concerns, it is an important indicator because it has implications for both present and future health. Studies, for example, have demonstrated that overweight youth are at increased risk for being overweight adults (Dipietro, Mossberg and Stunkard, 1994; Kvaavik, Tell and Klepp, 2003; Daniels, 2006).
FITNESSGRAM offers two different alternatives for body composition (body fat estimates and BMI).

Body fat can be estimated with a number of methods but the most common approaches in schools are the use of skinfold calipers or portable bioelectric impedance analyzers (BIA). Skinfold measurements are calculated by measuring the skinfold thickness in the triceps and calf. Skinfold measurements are the recommended measure of body composition in the FITNESSGRAM battery since they are easy to use and highly correlated to total body fatness. Each measurement should be taken a total of three times on the right hand side and then averaged. With skinfolds it is important to practice prior to collecting data, to gain familiarity and consistency within the process. The formula FITNESSGRAM uses to calculate percent of body fat from skinfold measurements was developed by Slaughter and Lohman (Slaughter, Lohman, Boileau, Horswill, Stillman, Vanloan and Bemben, 1988; Meredith and Welk, 2010). The formula states:

\[
\text{Body Fat Percentage} = 1.51 \times \text{BMI} - 0.70 \times \text{age} - 3.6 \times \text{sex} + 1.4 \quad (R^2 = 0.38, \text{SE of estimate (SEE) 4.4% BF%})
\]

BIA devices provide a simple way to estimate total body fat percentage. These devices calculate the fat percentage by sending a current through the body that measures the body’s resistance to that particular current. A body that has more fat will have greater resistance to the current, while a body that has more muscle will have less resistance to the current. There are a variety of BIA devices available but the most common devices for schools are simple field-based devices (Meredith and Welk, 2010). One type requires participants to stand on it with bare feet, while the other requires participants to use a handgrip system in which participants grip the handles of the device while extending the
arms. With both units a small current, that is not noticeable, is sent through the body and the resistance to current flow is then used to estimate body water (and then indirectly, body fatness). Internal predication equations in these devices are used to generate the estimate of body fatness.

BMI is an alternative indicator of body composition but it does not reflect the actual degree of body fatness. BMI is simply an indicator of mass relative to height and is calculated as mass (kg) / height (m$^2$). The limitation of BMI is that a person can have a larger muscle mass and have a high BMI without being over fat. Despite this limitation, BMI is widely used in public health research and is an important indicator for school based fitness evaluations (Meredith and Welk, 2010).

Studies have examined the reliability and validity of the three body composition measures (BIA, skinfold measurements and BMI). A study conducted by Gutin et al. (1996) used intra-class correlation coefficients (ICC), Bland-Altman plots, and Spearman rank correlations to examine test-retest reliability. Gutin et. al. (1996) compared measures of body fat percentages on 9-11 year old children, using skinfold measurements and BIA and used dual-energy X-ray absorptiometry (DXA) as the standard measure. The test-retest reliabilities for all three methods displayed ICC > 0.994, indicating no significant differences between trials. After calculating the percent of body fat all Spearman r values > 0.83, the three methods were found to be highly correlated.

While body fatness and BMI represent different indicators of body composition, studies show that they are moderately to highly correlated with each other (Pietrobelli, Faith et al., 1998). Differences in BMI and BIA outcomes have also been shown to be explained when height and weight are accounted for (Jackson, Pollock et al., 1988). Lukaski,
Bolonchuk, Hall, and Siders (1986) examined the validity and reliability of the bioelectrical impedance approach. For both males and females, densitometrically determined fat-free mass was compared to fat-free mass using the bioelectrical impedance. Both males and females were highly correlated, $r = 0.979$ and $r = 0.954$, respectively. In a study conducted by Heitmann (1990), BMI, skinfolds, and bioelectrical impedance were all found to provide reliable body fat estimates. When comparing the estimates of body fat by using multiple regression equations between 1) BIA 2) BMI and 3) skinfold, none of the differences found were significantly different from zero. The reference was determined by using a four-compartment-model based on measurements of total body water and potassium. Cable et al (2001) and Jebb et al (2000) conducted studies comparing BIA to criterion measures (dual energy X-ray absorptiometry and under-water weighing), and no significant differences were found. Utter et al (2001) provided research suggesting that bioelectrical impedance accurately predicts body fat percentage when compared to skinfolds.

When measuring body composition it is important to think about the psychological affect it may have in youth. Many studies have been conducted to look at the relationship between body composition and youth’s self-perceptions (Raustorp, Pangrazi and Stahle, 2004; Crocker, Sabiston, Kowalski, McDonough and Kowalski, 2006; Duncan, Al-Nakeeb, Nevill and Jones, 2006; Lubans and Cliff, 2011). Lubbans and Cliff (2011) found perceived body attractiveness and adiposity levels to be predictors of physical self-worth in females. Others have shown that the females that the relationship between BMI and self-perception tends to be stronger in females (Raustorp, Pangrazi et al., 2004; Duncan, Al-Nakeeb et al., 2006). Self-perceptions are thought to have important effects on overall self-esteem and on
adoption of health behaviors. Crocker et al. (2006) specifically found that self-perceptions can impact and predict a specific behavior such as physical activity.

**FITNESSGRAM Criterion-Referenced Standards**

A unique aspect of FITNESSGRAM is that the fitness tests are evaluated using criterion-referenced standards that are based on how much fitness is needed for good health (The Cooper Institute, 2011). Most fitness batteries, including the President’s Council on Physical Fitness and Sports, use normative standards that only provide relative indicators of fitness achievement. Normative standards provide an understanding and comparison of how their fitness levels measure up to those of an age matched reference population. Although norm-referenced standards are easy to calculate as long as a representative sample is available, there are three major limitations. These limitations include a) cost constraints associated with regularly updating the standards, b) the sample that were used to create the referenced population (i.e. if used an above average group as supposed to average), and finally c) the tendency to reward youth who are already fit, whereas possible discouragement may happen for those that don’t meet the standard.

Criterion-referenced standards provide a more effective way to evaluate individual and group fitness status. Criterion-referenced standards base feedback on whether a child has sufficient fitness for good health. This de-emphasizes individual comparisons and enables youth to set individual goals rather than performance goals. The original FITNESSGRAM standards used two different zones that divided the results from each fitness test into either the healthy fitness zone (HFZ) or the needs improvement zone (NI). While these standards served the program well, a limitation is that there isn’t a major difference between youth that fall slightly above or below the threshold.
New FITNESSGRAM standards were recently released that categorize fitness into three zones (Meredith and Welk, 2010). The use of three zones makes it possible to provide more effective and prescriptive messages to youth. Individuals falling above the criterion-referenced standard are classified in the HFZ, while all of those falling below the standard will be classified in the NI zone. The NI zone is subdivided into two categories, labeled NI – low risk or NI – high risk. Those in the NI zone are provided with information of risks associated with low fitness, as well as ways to get into the HFZ. While the standards and scoring procedures have changed, the tests are still implemented in the same way (The Cooper Institute, 2011). The sections below will provide summaries of the new standards and compare the new values with the old ones. Overall, there were more changes for females than males. Reasoning is provided along with the description of the new standards. In general, fewer females are likely to meet the new HFZ standards compared to the previous standards but the relationships may vary considerably by age or grade (Welk, Going et al., 2011).

**Development of the New Body Composition Standards**

Maintaining a healthy adiposity level and receiving early identification is key in helping youth reduce health risks. The new body composition standards were developed using nationally representative data from the National Health and Examination Survey (NHANES). The database enabled levels of BMI and body fatness to be related directly to the presence of metabolic syndrome, a group of risk factors that contribute to a variety of diseases including, coronary heart disease, type 2 diabetes, and stroke (Ford, Giles and Dietz, 2002). A unique feature of the new standards is that they also take into account differences in growth and maturation by gender and by age. The creation of these standards required several steps. Researchers first established least mean square (LMS) curves for body fatness
to define normal changes with growth across different ages (Laurson, Eisenmann and Welk, 2011c). The procedures are similar to those used to create the Centers for Disease Control and Prevention (CDC) growth charts for BMI (Centers for Disease Control and Prevention, 2010), but in this case, they reflect the normal developmental pattern for body fatness in males and females, not at risk for metabolic syndrome.

The actual thresholds and associated fitness zones were then established using receiver operator characteristic (ROC) curves that allow sensitivity and specificity to be compared for all possible thresholds (Laurson, Eisenmann and Welk, 2011b). A high threshold was established by emphasizing specificity and a lower one was established by emphasizing sensitivity. This also creates a third zone between the two threshold values. Youth with fat levels below the bottom standard who achieve the HFZ are considered at low risk for metabolic syndrome while youth with values that exceed the higher values are at a clear risk for metabolic syndrome. Youth that fall in between the two lines have a moderate risk. Figure 1 through 4 display the new standards in comparison with the old standards for both males and females in both BMI and body fat percentages (Welk, Maduro, Laurson and Brown, 2011). A final step in the process involved equating body fat values to BMI values. This was done using another ROC procedure that linked body fat to values of BMI (Laurson, Eisenmann et al., 2011b). The resulting standards for BMI are criterion-referenced health standards but it is useful to note that the values correspond fairly closely to the widely adopted CDC percentile standards used to define overweight and obesity.

The new body composition standards provide health standards that reflect potential risk for metabolic syndrome. The standards are slightly different than the previous ones used with the FITNESSGRAM program and this will lead to differences in the percentages of
youth that achieve the healthy fitness zones. Descriptive information is needed to understand how the new standards will change classifications of youth. Comparisons of the old and new standards are provided below.

**Development of the New Aerobic Capacity Standards**

The aerobic capacity standards were developed using the same approach used for the body composition. Developmental curves were used to establish growth trends for aerobic fitness (Eisenmann, Laurson and Welk, 2011) and the final standards were set using ROC curves (Welk, Going et al., 2011). Figure 5 and 6 display the new standards in comparison with the old standards for both males and females in aerobic capacity (Welk, Maduro et al., 2011). The new standards are slightly easier for males and slightly harder for females compared with the previous standards (Welk, Going et al., 2011). Therefore, fewer females and more males are likely to achieve the HFZ.

Special consideration was given to try to improve classification between the two assessments. The previous standards were based on estimated aerobic capacity (VO$_2$ max) but different approaches and equations were used for the estimation. Aerobic capacity in the one-mile run was estimated using the equation by Cureton et al (1995), while aerobic capacity in the PACER was estimated using an equation by Leger et al (1988). Both equations yielded reasonable estimates but some discrepancies in achievement rates had been previously noted (Welk, Meredith et al., 2010). The discrepancies were attributed to the fact that the equation used for the one-mile run included a BMI term but the PACER equation did not. The inclusion of BMI is considered important for accurate estimation, therefore a test equating approach was used to promote better classification agreement. With the new standards, PACER scores are first converted into estimated one-mile run times using
validated algorithms (Zhu, Plowman and Park, 2010). The values are then processed using the same prediction equation used to estimate aerobic capacity for the one-mile run. A cross validation study by Boiarskaia et al. (2011) determined that there were no significant differences between actual VO$_2$ max scores, and VO$_2$ max estimates calculated from the one-mile run or the from the PACER test. The use of a common outcome measure and the use of the same equation for estimating aerobic capacity provided a more standardized way to generate aerobic capacity estimates. However, there are some challenges with the new standards. Teachers must now collect and record height and weight with all fitness assessments since the one-mile run prediction equation includes a BMI term. Teachers may also not be able to provide students with target one-mile run times or laps needed to achieve the HFZ. Research is needed to further examine the utility of this test equating approach and the impact that the new standards have on fitness achievement in youth.

**Evaluating Measurement Agreement**

The new standards provide a more defensible way to evaluate fitness and fatness in youth. A recent study evaluated the classification of fitness with the old standards and the new ones (Welk, Maduro et al., 2011). The study also examined classification agreement between the different measures of fitness – for both body composition (BMI and body fat) and aerobic capacity (PACER and one-mile run). The researchers reported improved classification agreement with the new method linking PACER scores to the one-mile run. However, worse classification was noted for the body composition comparison. The researchers expected that the procedures used to match BMI values to body fat in the new standards would lead to improved classification agreement but the results did not indicate this (Welk, Maduro et al., 2011). A possible reason for the discrepancy was the use of a
portable BIA device for estimating body fatness. Additional research is needed to test fitness classification and classification agreement.

The primary goal in the present study was to examine measurement and classification agreement between alternative tests of aerobic capacity and body composition. A secondary goal was to measure agreement between males and females of different ages.
CHAPTER III: METHODS

This project was part of a larger study through the Iowa FITNESSGRAM Initiative – a participatory-action research partnership that provides schools throughout Iowa with support and training in the use of the FITNESSGRAM program. As partners in the research, schools agreed to conduct FITNESSGRAM testing as part of their normal physical education programming and share their data with our research group. This research was viewed as exempt by the Iowa State University Institutional Review Board (IRB).

Participating Schools
A total of 78 schools are currently enrolled in the Iowa FITNESSGRAM partnership. The present study used data from two schools in Iowa; a high school aged 14 – 18 years old (grades 9 – 12) and a junior high school aged 12 – 14 years old (grades 7, 8). Schools agreed to let researchers administer aerobic capacity and body composition tests.

Instruments
The study examined the measurement agreement and classification agreement between alternative assessments of aerobic capacity and body composition in the FITNESSGRAM fitness battery. The FITNESSGRAM program offers two different types of tests to determine a child’s level of aerobic capacity, the PACER test and the one-mile run. FITNESSGRAM also offers two different alternatives for body composition (body fat estimates and BMI). Detailed descriptions of the alternative assessments are provided below.

Aerobic Capacity Assessments
PACER Test: The PACER test was adapted from the 20-meter shuttle run (Leger, Mercier et al., 1988), refer to Appendix A. The test is set at a time cadence, in which it
begins at an easy steady pace, and gradually gets more difficult each minute as time progresses. The PACER test is the recommended test in the FITNESSGRAM battery since it has a number of important advantages (Meredith and Welk, 2010). Studies have found the PACER test to be a reliable and valid measure of aerobic capacity (Leger, Mercier et al., 1988; Liu, Plowman et al., 1992; Plowman and Liu, 1999; Mahar, Guerieri et al., 2011).

**One-Mile Run:** The one-mile run test also measures aerobic capacity and is an alternative to the PACER test, refer to Appendix B. The objective of this test is to run one-mile as fast as possible. If a student is not able to run the entire distance, walking is permitted. George et al. (1993) found the one-mile run to produce an accurate estimation of aerobic capacity.

**Body Composition Assessments**

**Body Fat Percent:** Skinfold measurements are the recommended measure of body composition in the FITNESSGRAM battery since they are easy to use and highly correlated to total body fatness (Meredith and Welk, 2010), refer to Appendix C. In the present study, Scanny Scientific Skinfold Calipers (Ann Arbor, MI, USA) were used. Skinfold measurements are calculated by measuring the skinfold thickness in the triceps and calf. The formula FITNESSGRAM uses to calculate percent of body fat from skinfold measurements was developed by Slaughter and Lohman (Slaughter, Lohman et al., 1988; Meredith and Welk, 2010).

An estimate of body fatness can also be obtained from portable BIA devices, refer to Appendix D. In the present study, an Omron handheld BIA device (Bannockburn, IL, USA) was used to provide an additional comparison measure. Skinfold measurements and BIA devices have been found to produce a reliable and valid measure of body fat percentage.
Body Mass Index (BMI): BMI is an alternative indicator of body composition but it does not reflect the actual degree of body fatness. BMI is simply an indicator of weight relative to height and is calculated as weight (kg) / height (m$^2$), refer to Appendix E. Digital scales (Lifesource MD Profit, Milpitas, CA, USA) and stadiometers (SECA Road Rod, Hanover, MD, USA) were used for measurements of weight and height. Heitmann (1990) and Pietrobelli et al. (1998) have both found BMI to be a valid measure of total body fatness.

Data Collection Procedures
The participating schools in the Iowa FITNESSGRAM Initiative were provided with detailed instructions on conducting fitness testing with the FITNESSGRAM battery. Training included access to online training resources, a copy of a training DVD, and email support. Schools were asked to complete all FITNESSGRAM assessments but were given choices about how they wanted to complete and report results in their school. The majority of schools elected to complete single assessments of aerobic capacity and body composition but the participating schools in the present project agreed to complete both aerobic fitness tests (PACER and the one-mile run), the BMI assessment, and both assessments of body fat (skinfold measurements and BIA). To ensure accurate data collection trained researchers were available to assist the schools in the administration of fitness tests and collection of the data.

Researchers entered fitness data into the program so that it could be aggregated through a customized (online) data aggregation utility. The program uses school ID numbers.
so it is possible to also link the child’s data to other school-level information including age, grade, gender, and ethnicity. The data aggregation utility however, enables data to be shared in a de-identified format to protect confidentiality.

Since the overall purpose of the study was to assess measurement and classification agreement between tests, it was important that the tests were conducted within a specific time period of one another. The PACER test and the one-mile run could not be conducted on the same day due to time restraints and fatigue, therefore these two tests were administered within ten days of one another. The order of administration was taken into consideration by assigning certain classes to complete the PACER first and other classes to complete the one-mile run first (counterbalanced test administration). All body composition assessments were collected on the same day. Trained skinfold caliper administrators practiced on youth as part of an after school program. Administrators were then tested against an expert prior to data collection.

Each skinfold measurement was taken a total of three times on the right hand side and then averaged. Conducting the tests within the same time frame helped to control for any daily variations in body composition. Since we were solely testing the relationships and classification agreement between the tests, each test was only conducted once. However, teachers involved in the Iowa FITNESSGRAM Initiative were encouraged to have students practice a test before it was formally conducted for the reporting. Since these schools used FITNESSGRAM as part of their curriculum, students had prior experience with the fitness tests. This familiarization process helps to ensure that students understand the nature of the assessments and have confidence in their ability to complete the assessments.
The FITNESSGRAM program is a user friendly program that allows teachers to quickly input the student’s information. The final data entered into the FITNESSGRAM version 9.2 reflected the individual scores obtained for each assessment along with height and weight. Built in algorithms are used to compute fitness scores, which are then related to the published HFZ standards based on the child’s age and gender. The individual results generated by the program reports the zones achieved by the children on the various assessments. Associated reporting utilities provided detailed breakdowns of the percentage of youth that achieve the various standards. De-identified data files were exported from this database tool and used in the subsequent statistical analyses.

**Statistical Analysis**

Data were combined across grades to create grade ranges that correspond to junior high school (grades 7, 8), and high school (grades 9 – 12). Descriptive analyses were conducted to examine differences in aerobic capacity and body composition levels by gender and grade level. Separate two-way ANOVA analyses were used to evaluate school type and gender differences in fitness levels with both sets of indicators (4 total ANOVA analyses).

The primary analyses focused on classification agreement between the two assessments of aerobic capacity and three assessments of body composition. The estimates of aerobic capacity and body composition from the different assessments were coded into one of the three fitness zones based on the new standards. Percent agreement was computed between the assessments and kappa statistics were used to understand the practical differences in agreement. Kappa was classified as poor if less than 0.20; fair agreement if between 0.20-0.40; moderate agreement if between 0.40-0.60; good agreement if 0.60-0.80
and very good agreement if 0.80-1.00. All analyses were performed using the Statistical Package for Social Sciences (SPSS) software version 17.0 (Chicago, IL, USA).
CHAPTER IV: RESULTS

Aerobic Capacity

Descriptive Statistics
The final sample of individuals participating in both measures of aerobic capacity was 701 students (334 males, 367 females). Results were analyzed separately for junior high school students in grades 7 and 8 (School Type 1: Mean age = 12.87, n = 300) and high school students in grades 9 – 12 (School Type 2: Mean age = 15.56, n = 401). Descriptive statistics representing the distribution frequency for aerobic capacity among gender and grade level as well as a breakdown of the means and standard deviations for both measures of aerobic capacity are included in Table 1. An outlier analysis was conducted to identify possible outliers in the data set for both aerobic capacity and body composition. Seven values exceeded the mean value by more than four standard deviations and were removed from the data set as they were skewing the results.

Evaluation of Measurement Agreement
Pearson correlations were used to determine the overall strength of the relationship between measures. The overall correlation between predicted aerobic capacity using the PACER and predicted aerobic capacity using the one-mile run was high (r = 0.94). The relationships were consistent when results were examined by gender (r males = 0.94, r females = 0.91) and school type (r school type 1 = 0.92, r school type 2 = 0.95). A 2-way (gender x school type) ANOVA was conducted to test for mean differences in aerobic capacity estimates. The overall F test was not statistically significant [F (3, 697) = 0.445, p > 0.05]. The main effect was not significant for gender (F = 0.079, p = 0.779), school type (F = 0.712, p = 0.399), or the interaction (F = 0.556, p = 0.456). The mean differences between
measures of aerobic capacity are graphed in Figure 7. A Bland-Altman plot showing the
difference between predicted levels of aerobic capacity against the means is provided in
Figure 8. The results did not indicate any clear systematic bias.

**Evaluation of Classification Agreement**

The classification agreement looked at the consistency with which the tests classified
youth into the different fitness zones. The breakdown of each zone and the specific
percentage by school type and gender is located in Table 2. The percent of students achieving
the HFZ was similar for the one-mile run (69.6%) and the PACER (72.6%). The percent of
students achieving the HFZ was also similar when examined by school type. For the junior
high school the percent of students achieving the HFZ was 74% for the one-mile run and
77% for the PACER. For the high school the percent of students achieving the HFZ was
66.1% for the one-mile run and 69.8% for the PACER.

The classification matrix for the combined sample is provided in Table 3. The total
percent classified similarly by the two tests (see values on the diagonal) was 92.7%. The
classification agreement was high when examined separately by school type and gender. In
the junior high school, 94.4% of the males and 88.3% of females were classified consistently.
In the high school, 93.7% of males and 93.4% of females were classified consistently. The
majority of misclassifications (5.8%) occurred when the PACER classified the student in a
higher fitness level than the one-mile run. Four percent of the sample was classified in the
HFZ by the PACER and in the NI – low risk zone by the one-mile run. Nearly two percent of
the sample was classified as NI – low risk by the PACER and NI – high risk by the one-mile
run.
Kappa statistics provided a statistical measure of classification agreement protecting against agreement by chance and can be seen in Table 4. These analyses were run for the combined sample as well as separately by gender and school type. Kappa for most measures was considered to have very good agreement (kappa all together = 0.84). The kappa score was relatively similar for the junior high school (0.84) and high school (0.87) males. The kappa score was slightly different for the junior high school (0.75) and high school (0.86) females.

**Body Composition**

**Descriptive Statistics**

The final sample of individuals participating in all three measures of body composition was collected on 889 students (412 males, 477 females). Results were analyzed separately for junior high school students in grades 7 and 8 (School Type 1: Mean age = 12.87, n = 398) and high school students in grades 9 – 12 (School Type 2: Mean age = 15.66, n = 491). Descriptive statistics representing the distribution frequency for body composition among gender and grade level as well as a breakdown of performance for all three measures of body composition are included are Table 5.

**Evaluation of Measurement Agreement**

Correlations between predicted percent of body fat using skinfolds and predicted percent of body fat using BIA were found to be positive and high for the combined sample (r = 0.86) as well as when examined by gender (r males = 0.86, r females = 0.82) and school type (r school type 1 = 0.84, r school type 2 = 0.87). A 2-way ANOVA was conducted to test for mean differences between gender and school type. The overall F test was significant [F (3,885) = 7.020, p < 0.01]. The main effect was not significant for gender (F = 2.303, p =
0.129), but it was for school type (F = 10.487, p = 0.001). The interaction term was also significant (F = 5.590, p = 0.018) indicating differential relationship for various age and gender combinations. Examination of the subgroups showed that the difference in estimates (skinfold - BIA) tended to be small and negative for the younger males, younger females, and older males but larger (> 1 unit of % body fat) and positive in older females (See Figure 9). A Bland-Altman plot showing the difference between predicted percent of body fat against the means is provided in Figure 10. The results indicated no systematic bias and good overall agreement with the differences between the measures of percent of body fat being around zero.

**Evaluation of Classification Agreement**

The classification agreement looked at the consistency with which the tests classified youth into the different fitness zones. The breakdown of each zone and the specific percentage by school type and gender is located in Table 6. The percent of students achieving the HFZ was relatively similar for BIA (58.8%), skinfold (61.3%), and BMI (55.2%), with the percent of students being classified in the HFZ slightly higher with skinfold, and slightly lower with BMI. The relationships were similar for both the junior high school sample (BIA = 57%, skinfold = 64%, and BMI = 55%) and high school sample (BIA = 61%, skinfold = 59%, and BMI = 55%).

The classification matrix for the combined sample is provided in Table 7. The total percent classified similarly by skinfold and BIA (see values on the diagonal) was 76.2%. The total percent classified similarly by skinfold and BMI (see values on the diagonal) was 68%. Finally, the total percent classified similarly by BIA and BMI (see values on the diagonal) was 67.7%. The majority of misclassifications appear to be when comparing skinfold and
BIA to BMI. The most common type of misclassification is with BMI values placing youth into the NI – high risk category while the skinfold or BIA placed students into the NI – low risk category. Looking at these cells, the rate of misclassification was 16.1% for BIA/BMI and 11.7% for skinfold/BMI. This was apparent regardless of gender or school type. A similar misclassification occurred with skinfold placing students into the HFZ and BIA/BMI placing them into NI – low risk. Looking at these cells, misclassification was 8.2% for both skinfold/BIA and skinfold/BMI.

Kappa statistics provided a statistical measure of classification agreement protecting against agreement by chance and can be seen in Table 8. These analyses were run for the combined sample as well as separately by gender and school type. Looking at the agreement between skinfold and BIA, kappa was considered to have moderate agreement (kappa all together = 0.575). Looking at the agreement between skinfold and BMI kappa was considered to have fair to moderate agreement (kappa all together = 0.454). Finally, looking at the agreement between BIA and BMI kappa was considered to have fair to moderate agreement (kappa all together = 0.462).
CHAPTER V: DISCUSSION

There were two goals for this study. The primary goal was to determine the overall agreement between alternative tests of aerobic capacity (PACER and the one-mile run) and between alternative tests of body composition (BIA, skinfold measurements and BMI). The secondary purpose was to determine if there were any gender or age differences associated with these classification agreements. The results generally supported the classification agreement between tests of aerobic capacity but agreement was lower among the body composition assessments. There were very few differences between school type and gender for measures of aerobic capacity, with more differences associated with school type and gender for measures of body composition.

Discussion of Aerobic Capacity Findings

The new FITNESSGRAM standards were designed to improve measurement agreement and classification agreement between the PACER and the one-mile run. A test-equating procedure is used to convert PACER scores into one-mile run times so that both the PACER and the one-mile run can use the same equation for estimation of aerobic capacity (Welk, Laurson, Eisenmann and Cureton, 2011). The utility of this approach was supported in a cross validation study (Boiarskaia, Boscolo et al., 2011). The results of the present study also support the utility of this new approach. The correlations observed between the two measurements of aerobic capacity were considerably higher than previous studies have reported with it remaining fairly consistent when examined separately by school type and gender (Dinschel, 1994; Mahar, Rowe, Parker, Mahar, Dawson and Holt, 1997). Non-significant differences were also reported between the estimates of aerobic capacity from the
two assessments. There was little or no evidence of any systematic bias in the plots, but the differences tended to be greater at higher average of aerobic capacities (see Figure 8). The PACER test tended to yield slightly higher estimates of aerobic capacity compared with the one-mile run as the average difference (PACER – one-mile run) was consistently below zero. However, these differences were not statistically significant.

The percentage of students achieving the HFZ was very similar between the two tests (with values differing by less than three percentage points across the various subgroups). A previous study using a large state-wide dataset revealed significant differences in the fitness distributions based on the one-mile run and those based on the PACER (Welk, Meredith et al., 2010). This discrepancy challenged the scientific basis of the FITNESSGRAM program since the two assessments are viewed as being interchangeable with each other. The present results indicate that the test-equating procedure does lead to improved classification agreement.

In addition to good overall measurement agreement between the PACER and the one-mile run there was good classification agreement. The overall classification was good at 92.7%. This is slightly higher than values reported using the older FITNESSGRAM standards (Mahar, Rowe et al., 1997). The kappa scores revealed slightly better agreement for males compared to females. Boiarskaiat et al. (2011) also found classification accuracy to be slightly higher for males compared to females. It is not clear if this is due to the standards or due to the nature of the tests. Although students were instructed to give it their all for both tests, discrepancies in the study may be due to inconsistent efforts given between tests. Overall the tests seem to show a high sense of agreement. These findings support to use of the PACER and one-mile run as alternative measures used for fitness testing.
The new standards were expected to change the distribution of fitness achievement between males and females. With the previous standards, a higher percentage of females compared to males achieved the HFZ. The revised standards made it more difficult for young females to achieve the HFZ but easier for young males to achieve them. In the present study we observed a clear tendency for higher fitness achievement among junior high school students compared to high school students but this pattern varied between gender. A higher number of junior high school males were reaching the HFZ compared to females; however, this was not the case for the high school. In the high school, slightly more females than males were reaching the HFZ. Welk, Maduro et al. (2011) found similar results to this study, with larger declines in the percentage of males reaching the HFZ as age increased and smaller age-related declines in females. It is not possible to determine what the true distribution of health risks are in the sample so future studies may be needed to examine the utility of the specific standards for detecting risk of disease. Research suggests that absolute and relative peak VO$_2$ remains stable throughout adolescence in males, while absolute and relative peak VO$_2$ appears to decrease from childhood to adolescence in females (Armstrong and Welsman, 1994; Eisenmann and Malina, 2002). The different patterns of growth during adolescence make this a challenging area of research. A study conducted by Ihaz et al. (2006) suggested a modified version of the PACER test may be more beneficial for overweight youth due to the importance of speed; however, this resulted in a less accurate prediction of VO$_2$. 
Discussion of Body Composition Findings

New body composition standards were developed to provide a way for teachers to identify students that may be at risk for health problems such as diabetes. The growth chart standards used by the Centers for Disease Control and Prevention (CDC) are widely used but they are based on percentiles rather than a health criterion. The FITNESSGRAM standards linked body fatness levels to potential risk for metabolic syndrome (Laurson, Eisenmann et al., 2011c). Once the body fat standards were created the researchers equated the body fat values to BMI values (Laurson, Eisenmann and Welk, 2011a). While the procedures are sound, previous research showed some discrepancies in classification agreement between the classifications from BIA-based body fat estimates versus those from BMI (Welk, Maduro et al., 2011). The present study evaluated this in more detail by utilizing two different measures of body fatness (skinfolds and BIA) but similar relationships were found.

Specific comparisons were made between skinfold and BIA to determine if they provided comparable estimates of body fat in this population. However, emphasis was placed on fitness classification and classification agreement since all three tests can determine what zone reflects the individual’s overall health for body composition. Overall the correlations between skinfold and BIA were consistently high for all comparisons indicating that they are related to each other. Previous research has found the two to be highly correlated as well (Lukaski, Bolonchuk et al., 1986; Utter, Scott et al., 2001). Minimal differences were found between the difference in skinfold and BIA in junior high school youth, however this was not the case for high school females. While junior high school males and females and high school males had negative differences (skinfold under predicted compared to BIA), high school females had positive differences (skinfold over predicted compared to BIA) (see
Figure 9). This is an interesting point as this has not been reported in previous studies. The reasoning behind this could be contributed to a variety of factors including inconsistencies with BIA equipment and the researchers’ limited access to practice on youth with a higher percent of body fat prior to data collection. The skinfold measurement administrators and instruments were consistent for both school types. In a study conducted by Jackson et al. (1988) height and weight was thought to account for the most variance in the BIA equation.

Looking at the Bland-Altman plot (see Figure 10) it does appear that as the average percent of body fat increased, the skinfold measurements were more likely to over predict as compared to BIA. BIA has been associated with a lower predictive error and standard error in comparison to standard anthropometric techniques such as skinfold measurements (Lukaski, Bolonchuk et al., 1986). Heitmann (1990) also found there to be lower residual error among BIA in comparison with the regression equations used for skinfold measurements.

The percent of students being classified in the different zones can help us understand specific school type and gender differences. The overall percentage of students being classified was not similar and this was most apparent between skinfold and BMI. Overall skinfold measurements (see Table 6) classified the most amount of students in the HFZ and the second most in NI – high risk, BIA classified the most in NI – low risk and the least in NI – high risk, and BMI classified the most as NI – high risk and the least in the HFZ.

When comparing tests (see Table 7), there were two main misclassification issues. The first was between skinfold and BIA as well as skinfold and BMI. Skinfold classified cases in the HFZ, while BIA and BMI would classify these cases in the NI – low risk zone. However, there were larger discrepancies when comparing skinfold to BMI and BIA to BMI. In a number of cases, skinfold and BIA classified individuals as NI – low risk but the BMI
value classified them as NI – high risk. This was apparent regardless of gender or school type. While BMI has been found to provide a reasonable proxy measure for assessing body composition the limitations have been well described. Previous research has suggested being cautious when comparing BMI to percent of body fat (Pietrobelli, Faith et al., 1998). There have also been higher variances associated with body fat estimates using BMI in comparison to BIA or skinfolds (Heitmann, 1990).

Emphasis in FITNESSGRAM is on classification into fitness zones so the comparisons of classification agreement are most important. Overall kappa statistics were highest when comparing skinfold and BIA, second strongest between BIA and BMI, and lowest of the three pairwise comparisons with skinfold and BMI. Overall the tests seem to show a moderate sense of agreement but less than would be expected considering the rigor used in establishing the BMI standards based on the body fat values (Laurson, Eisenmann et al., 2011a). These findings support the use of BIA, skinfold measurements, and BMI as alternative measures used for fitness testing; however, based on the results, there are still considerable inconsistencies in classification.

Discrepancies between measures of body composition can be due to a number of factors including measurement error. Although test administrators were consistent throughout data collection, practiced prior to data collection, and were tested against an expert, there may still be measurement error due to inexperience. However, researchers were likely to be more experienced with skinfold caliper use than the average physical education teacher, the targeted administrator for these tests. If an untrained physical education teacher administered these tests the potential for measurement error would be even greater. A reason for the discrepancies between skinfold and BIA against BMI categorizing some cases into the HFZ,
primarily athletes, is that BMI may erroneously classify students in the NI – low risk zone if their BMI (weight relative to height) is higher due to a larger muscle mass. Lean students with high muscle mass would tend to have slightly higher BMI and low body fat values.

The percentage of students being classified in the HFZ appears to be lower with the new standards in comparison to the previous standards with approximately 20-30% of students being classified as NI (Welk, Maduro et al., 2011). However, the results of body composition in this study agree with a previous study looking at the classification agreement. In this study, more individuals were classified in the HFZ by percent of body fat then by BMI. Previous research has also reported lower classification agreement and kappa scores among the comparison between BIA and BMI (Welk, Maduro et al., 2011). The results do not appear to be due to the use of BIA as skinfold values yielded the same pattern. Without a criterion measure it is not possible to determine the true relationships.

Since body composition has the potential to greatly affect a child’s self-perception and self-worth it is important to look at the potential affect these measurements and classifications would have on youth (Crocker, Sabiston et al., 2006; Wiersma and Sherman, 2008; Lubans and Cliff, 2011). Based on the results, fewer students would be labeled as “high risk” with percent of body fat than with BMI. From a psychological perspective, it would be preferable to mistakenly classify a student as NI – low risk when they should have been classified as NI – high risk, than to erroneously classify a student as NI – high risk if, in fact, they were actually just a heavily muscled student with a slightly higher body fat level and a NI-low risk rating of body fat. In this regard, BMI be the worst measurement of body composition as it has the potential to have strongest negative results or associations with physical activity and self-perception. Classifying more students as NI – low risk as opposed
to NI – high risk would have the most potential to motivate youth to be healthier and adapt lifelong physical activity habits.

**Summary of Results and Implications for Youth Fitness Testing**

After administering the various fitness tests, recommendations for future uses of FITNESSGRAM include using either measure of aerobic capacity (PACER or the one-mile run) and BIA for body composition. The tests of aerobic capacity had strong correlation as well as kappa statistics. The students did appear to enjoy the PACER test more, it is the recommended test for FITNESSGRAM and does include various benefits such as a built-in warm up, it teaches youth how to pace themselves, and those that have poorer performance finish first as opposed to last with the one-mile run (Meredith and Welk, 2010). The PACER test is also very easy to administer and can be done in a gym. There is both a 15- and 20-meter version available. For body composition measurements I would recommend the BIA device, since it was associated with the potential for least hurtful error. Although skinfold is the recommended assessment for body composition it has the potential to be associated with the highest measurement error if skinfold administrators are inexperienced or untrained. Since aerobic capacity measures now require a BMI term teachers may solely use BMI as their measure of body composition leading to potential hurtful error. The primary purpose of fitness testing is to provide students with feedback on their current level of fitness and ways to improve. While it is important to take the psychological affects into consideration it is also important to provide teachers with an accurate and feasible way to assess both aerobic capacity and body composition and is currently done with the use of FITNESSGRAM.

This study provides valuable information regarding the classification agreement among the new FITNESSGRAM standards. The results show that alternative tests can be
used in the FITNESSGRAM program to provide similar information about fitness and body composition. While the results with aerobic capacity revealed very good agreement the agreement among the body composition measures was weaker. Although some have questioned the purpose and the usefulness of fitness testing (Corbin, Pangrazi et al., 1995; Rowland, 1995; Cale, Harris et al., 2007), it is a common component of most physical education programs (Lee, Burgeson et al., 2007; Silverman, Keating et al., 2008; Welk, 2008). It can provide useful information for teachers, administrators and researchers but it is important to ensure that the alternative tests give similar information. Accurate fitness classification is also important to ensure that appropriate feedback is provided to children and parents.

One strength of the present study is the relatively large sample size (there were 701 students for aerobic capacity, and 889 students for body composition). The sample size was fairly balanced among different grade levels as well as between gender. Students and researchers followed the proper FITNESSGRAM protocol when conducting the different assessments. Students were familiar with the tests prior to data collection and were provided with an opportunity to practice. Although aerobic capacity assessments were not able to be conducted on the same day due to fatigue, they were conducted within 10 days of one another to ensure no changes in VO$_2$, plus counterbalance was taken into consideration to prevent any test bias. Body composition tests were all conducted on the same day to ensure accurate agreement. Body composition administrators and specific instruments remained consistent throughout data collection. Administrators not only practiced on youth prior to data collection but were compared against an expert. Teachers were actively involved in the test administration process to ensure students were fully participating in the assessments.
Data was entered into the online program by researchers to ensure accurate data entry and ID numbers were used to de-identify student’s data to ensure confidentiality. Students and parents were provided with individualized reports showing the students fitness results and personalized feedback.

Although there were many strengths displayed in this study there were multiple limitations as well. The sample was all collected within a predominately Caucasian, rural, upper-middle class school district. The sample is also not representative of all grade levels. While only junior high (7, 8) and high school were used (9 – 12), it would be nice to have a more representative sample including elementary and middle school (K – 6). The location of the one-mile run was different between the junior high school and high school. The tests were conducted during the winter, and while the high school had access to an indoor track the junior high school did not and ran outside. Since the purpose of this test was the classification agreement and any differences associated with age or gender, a golden standard was not used in comparison with measures of body composition. Although each measure has been proven to be reliable and accurate one measure may be more reliable in this situation, compared to another. Since the study was conducted during the school day the students did their aerobic capacity or body composition tests anytime from 8 am to 3 pm. Students may have different results based on the time of day their assessment was administered. Variations may be most notable in BIA, while lunch may play a role in the students’ hydration making the body fat percentage vary slightly.

This study provides multiple opportunities for future research. One important study would include taking a subsample of the population used for the present study and comparing their already collected body fat percentage (BIA and skinfold measurements) and BMI
against a criterion measure in the lab for example the Bod Pod or underwater weighing as conducted by Cable et al. (2001). Although body composition and aerobic capacity were used since they are closely related to a variety of chronic diseases including coronary heart disease, hypertension and type 2 diabetes mellitus (Anderssen, Cooper et al., 2007). FITNESSGRAM assesses other dimensions of fitness with multiple assessments. For example muscular strength and endurance can be assessed with either the curl-up, trunk lift, push-up, pull-up, or flexed arm hang. Since many males view muscular fitness and endurance as an important evaluation of physical self-worth (Lubans and Cliff, 2011) it would be important to continue conducting research to support measurement agreement and classification agreement for the other three dimensions (muscular endurance, muscular strength and flexibility). A third suggested study would be to examine the affect that being classified in the NI – high risk zone would have on the psychological well-being of youth and whether it is different among gender. The fourth and final suggestion for future research would be to conduct a similar study focusing on the aerobic capacity and body composition assessments, but including a more diverse sample. This could include more variety in age as well as a more diverse population.

**Conclusion**

Physical fitness testing is a common component of most physical education programs (Lee, Burgeson et al., 2007) and can provide teachers with a way to facilitate fitness education, deliver personalized feedback to each student, and provide students with a way to track their fitness levels over time (Welk and Meredith, 2008). The FITNESSGRAM program provides teachers with a variety of fitness tests for assessing the different dimensions of fitness in schools. A unique advantage of the battery is that the fitness tests
have good measurement properties and are scored using criterion-referenced standards (The Cooper Institute, 2011). Teachers have choices about what specific assessments they use but it is important to determine if the tests classify students similarly and provide related information about youth’s fitness levels. FITNESSGRAM's new standards reflect age- and gender- specific standards needed for good overall fitness (Meredith and Welk, 2010). While the PACER test is the recommended test to measure aerobic capacity and skinfold measurements are the recommended test to measure body composition alternative tests are often used. The findings of this study suggest that there is strong agreement between measures of aerobic capacity and moderate levels of agreement between measures of body composition.
REFERENCES

Anderssen, S. A., Cooper, A. R., Riddoch, C., Sardinha, L. B., Harro, M., Brage, S. and
Andersen, L. B. (2007). "Low cardiorespiratory fitness is a strong predictor for clustering of


of an Equating Method Linking Aerobic FITNESSGRAM (R) Field Tests." American

Cable, A., Nieman, D. C., Austin, M., Hogen, E. and Utter, A. C. (2001). "Validity of leg-to-
leg bioelectrical impedance measurement in males." Journal of Sports Medicine and Physical

Cale, L., Harris, J., Chen, M. H., Corbin, C., Fox, K., Morrow, J. R. and Plowman, S. A.
(2007). "More than 10 years after "The horse is dead...": Surely it must be time to

EXERCISE, AND PHYSICAL-FITNESS - DEFINITIONS AND DISTINCTIONS FOR


dismount"." Pediatric Exercise Science 7: 347-351.

Crocker, P. R. E., Sabiston, C. M., Kowalski, K. C., McDonough, M. H. and Kowalski, N.
(2006). "Longitudinal assessment of the relationship between physical self-concept and
health-related behavior and emotion in adolescent girls." Journal of Applied Sport
Psychology 18(3): 185-200.

"A GENERALIZED EQUATION FOR PREDICTION OF VO2PEAK FROM 1-MILE RUN


Figure 1 Body Composition: New standards compared to previous standards for males.
**Females BMI Standards**

![BMI Chart for Females](chart.png)

**Figure 2 Body Composition: New standards compared to previous standards for females.**
Figure 3 Body Composition: New standards compared to previous standards for males.
Females Body Fat Standards

![Graph showing body fat standards for females across different age groups. The graph compares new standards to previous standards, highlighting various thresholds for NI - High Risk, NI - Low Risk, Healthy Fitness Zone, and Very Lean.](image)

**Figure 4 Body Composition:** New standards compared to previous standards for females.
Figure 5  Aerobic Capacity: New standards compared to previous standards for males.
Females Aerobic Capacity Standards

Figure 6 Aerobic Capacity: New standards compared to previous standards for females.
Figure 7 Aerobic Capacity: Mean of the difference between the PACER minus the One Mile Run by school type and gender.
Figure 8: Aerobic Capacity: Association of the difference between measures of aerobic capacity against the mean aerobic capacity.
Mean of the Difference Between Measures of Percent of Body Fat (Skinfold - BIA)

Figure 9 Body Composition: Mean of the difference between Skinfold Measurements minus BIA by school type and gender.
Figure 10 Body Composition: Association of the difference between measures of percent of body fat against the mean percent of body fat.
### Table 1: Aerobic Capacity: Means and standard deviations for the PACER and One-Mile Run by grade and gender.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>n</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td></td>
<td>PACER Aerobic Capacity (ml/kg/min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>45.5 ± 5.6</td>
<td>88</td>
<td>41.2 ± 5.5</td>
</tr>
<tr>
<td>8</td>
<td>46.6 ± 5.8</td>
<td>74</td>
<td>41.7 ± 3.9</td>
</tr>
<tr>
<td>9</td>
<td>46.6 ± 6.9</td>
<td>37</td>
<td>40.9 ± 5.1</td>
</tr>
<tr>
<td>10</td>
<td>43.1 ± 6.4</td>
<td>33</td>
<td>39.4 ± 4.8</td>
</tr>
<tr>
<td>11</td>
<td>44.8 ± 8.0</td>
<td>56</td>
<td>39.4 ± 4.4</td>
</tr>
<tr>
<td>12</td>
<td>44.7 ± 8.0</td>
<td>46</td>
<td>39.2 ± 4.9</td>
</tr>
<tr>
<td>Total</td>
<td>334</td>
<td>367</td>
<td></td>
</tr>
<tr>
<td></td>
<td>One-Mile Run Aerobic Capacity (ml/kg/min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>45.9 ± 5.9</td>
<td>88</td>
<td>41.6 ± 5.9</td>
</tr>
<tr>
<td>8</td>
<td>47.3 ± 6.2</td>
<td>74</td>
<td>42.7 ± 4.5</td>
</tr>
<tr>
<td>9</td>
<td>48.1 ± 6.8</td>
<td>37</td>
<td>42.1 ± 5.7</td>
</tr>
<tr>
<td>10</td>
<td>43.5 ± 6.6</td>
<td>33</td>
<td>39.8 ± 5.3</td>
</tr>
<tr>
<td>11</td>
<td>45.7 ± 8.4</td>
<td>56</td>
<td>40.3 ± 5.2</td>
</tr>
<tr>
<td>12</td>
<td>45.2 ± 8.9</td>
<td>46</td>
<td>39.3 ± 5.1</td>
</tr>
<tr>
<td>Total</td>
<td>334</td>
<td>367</td>
<td></td>
</tr>
</tbody>
</table>

SD = Standard Deviation
Table 2 Aerobic Capacity: Percent of students classified in the different zones by school type and gender.

<table>
<thead>
<tr>
<th>Aerobic Capacity</th>
<th>School Type 1</th>
<th></th>
<th>School Type 2</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
<td>All Together</td>
<td>Males</td>
<td>Females</td>
<td>All Together</td>
<td>All Together</td>
</tr>
<tr>
<td>PACER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFZ</td>
<td>80.9%</td>
<td>72.5%</td>
<td>77.0%</td>
<td>67.4%</td>
<td>71.6%</td>
<td>69.8%</td>
<td>72.6%</td>
</tr>
<tr>
<td>NI - Low Risk</td>
<td>11.7%</td>
<td>12.3%</td>
<td>12.0%</td>
<td>9.9%</td>
<td>13.5%</td>
<td>12.0%</td>
<td>12.1%</td>
</tr>
<tr>
<td>NI - High Risk</td>
<td>7.4%</td>
<td>15.2%</td>
<td>11.0%</td>
<td>22.7%</td>
<td>14.8%</td>
<td>18.2%</td>
<td>15.3%</td>
</tr>
<tr>
<td>One-Mile Run</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFZ</td>
<td>79.6%</td>
<td>67.4%</td>
<td>74.0%</td>
<td>64.0%</td>
<td>67.7%</td>
<td>66.1%</td>
<td>69.6%</td>
</tr>
<tr>
<td>NI - Low Risk</td>
<td>10.5%</td>
<td>18.8%</td>
<td>14.3%</td>
<td>10.5%</td>
<td>17.0%</td>
<td>14.2%</td>
<td>14.2%</td>
</tr>
<tr>
<td>NI - High Risk</td>
<td>9.9%</td>
<td>13.8%</td>
<td>11.7%</td>
<td>25.6%</td>
<td>15.3%</td>
<td>19.7%</td>
<td>16.2%</td>
</tr>
</tbody>
</table>

HFZ = Health Fitness Zone
NI = Needs Improvement Zone
Table 3 Aerobic Capacity: Overall classification agreement for measures of aerobic capacity (One-Mile Run and PACER).

<table>
<thead>
<tr>
<th>PACER</th>
<th>HFZ</th>
<th>NI - Low Risk</th>
<th>NI - High Risk</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>HFZ</td>
<td>68.6%</td>
<td>4.1%</td>
<td>0.1%</td>
<td>92.7%</td>
</tr>
<tr>
<td>NI - Low Risk</td>
<td>0.6%</td>
<td>9.7%</td>
<td>1.7%</td>
<td></td>
</tr>
<tr>
<td>NI - High Risk</td>
<td>0.3%</td>
<td>0.4%</td>
<td>14.4%</td>
<td></td>
</tr>
</tbody>
</table>

HFZ = Health Fitness Zone

NI = Needs Improvement Zone
Table 4 Aerobic Capacity: Kappa agreement between measures of aerobic capacity (PACER and One-Mile Run) by school type and gender.

<table>
<thead>
<tr>
<th>Kappa of Aerobic Capacity</th>
<th>Males</th>
<th>Females</th>
<th>All Together</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Type 1</td>
<td>0.84</td>
<td>0.75</td>
<td>0.79</td>
</tr>
<tr>
<td>School Type 2</td>
<td>0.87</td>
<td>0.86</td>
<td>0.87</td>
</tr>
<tr>
<td>All Together</td>
<td>0.86</td>
<td>0.82</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Table 5 Body Composition: Means and standard deviations for BIA, Skinfold Measurements and BMI by grade and gender.

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>n</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>BIA (Percent)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>23.7 ± 8.2</td>
<td>114</td>
<td>26.3 ± 7.6</td>
</tr>
<tr>
<td>8</td>
<td>20.9 ± 8.8</td>
<td>87</td>
<td>25.3 ± 6.8</td>
</tr>
<tr>
<td>9</td>
<td>21.2 ± 10.1</td>
<td>39</td>
<td>26.7 ± 6.6</td>
</tr>
<tr>
<td>10</td>
<td>22.7 ± 9.4</td>
<td>42</td>
<td>28.3 ± 6.1</td>
</tr>
<tr>
<td>11</td>
<td>22.3 ± 10.6</td>
<td>76</td>
<td>27.6 ± 6.2</td>
</tr>
<tr>
<td>12</td>
<td>19.4 ± 8.7</td>
<td>54</td>
<td>26.6 ± 6.8</td>
</tr>
<tr>
<td>Total</td>
<td>412</td>
<td>477</td>
<td>889</td>
</tr>
<tr>
<td>Skinfold (Percent)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>22.5 ± 10.8</td>
<td>114</td>
<td>25.5 ± 8.6</td>
</tr>
<tr>
<td>8</td>
<td>21.6 ± 11.3</td>
<td>87</td>
<td>24.8 ± 8.1</td>
</tr>
<tr>
<td>9</td>
<td>20.8 ± 11.4</td>
<td>39</td>
<td>26.6 ± 8.0</td>
</tr>
<tr>
<td>10</td>
<td>21.5 ± 10.4</td>
<td>42</td>
<td>29.3 ± 9.0</td>
</tr>
<tr>
<td>11</td>
<td>22.7 ± 13.5</td>
<td>76</td>
<td>30.2 ± 9.3</td>
</tr>
<tr>
<td>12</td>
<td>19.9 ± 11.7</td>
<td>54</td>
<td>28.4 ± 10.2</td>
</tr>
<tr>
<td>Total</td>
<td>412</td>
<td>477</td>
<td>889</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>21.47 ± 4.3</td>
<td>114</td>
<td>21.9 ± 4.8</td>
</tr>
<tr>
<td>8</td>
<td>22.2 ± 4.9</td>
<td>87</td>
<td>22.0 ± 4.3</td>
</tr>
<tr>
<td>9</td>
<td>23.8 ± 6.0</td>
<td>39</td>
<td>23.1 ± 4.9</td>
</tr>
<tr>
<td>10</td>
<td>25.1 ± 6.1</td>
<td>42</td>
<td>24.2 ± 4.7</td>
</tr>
<tr>
<td>11</td>
<td>25.7 ± 7.0</td>
<td>76</td>
<td>24.3 ± 5.2</td>
</tr>
<tr>
<td>12</td>
<td>24.9 ± 6.0</td>
<td>54</td>
<td>24.1 ± 4.8</td>
</tr>
<tr>
<td>Total</td>
<td>412</td>
<td>477</td>
<td>889</td>
</tr>
</tbody>
</table>

SD = Standard Deviation
Table 6 Body Composition: Percent of students classified in the different zones by school type and gender.

<table>
<thead>
<tr>
<th>Body Composition</th>
<th>School Type 1</th>
<th>School Type 2</th>
<th>School Type 1</th>
<th>School Type 2</th>
<th>School Type 1</th>
<th>School Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
<td>All Together</td>
<td>Males</td>
<td>Females</td>
<td>All Together</td>
</tr>
<tr>
<td>BIA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very Lean</td>
<td>1.0%</td>
<td>2.5%</td>
<td>1.8%</td>
<td>2.4%</td>
<td>1.4%</td>
<td>1.8%</td>
</tr>
<tr>
<td>HFZ</td>
<td>55.7%</td>
<td>57.9%</td>
<td>56.8%</td>
<td>53.6%</td>
<td>65.7%</td>
<td>60.5%</td>
</tr>
<tr>
<td>NI - Low Risk</td>
<td>34.3%</td>
<td>21.5%</td>
<td>32.9%</td>
<td>27.5%</td>
<td>23.6%</td>
<td>25.3%</td>
</tr>
<tr>
<td>NI - High Risk</td>
<td>9.0%</td>
<td>8.1%</td>
<td>8.5%</td>
<td>16.6%</td>
<td>9.3%</td>
<td>12.4%</td>
</tr>
<tr>
<td>Skinfold</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very Lean</td>
<td>0.5%</td>
<td>1.5%</td>
<td>1.0%</td>
<td>0.5%</td>
<td>1.4%</td>
<td>1.0%</td>
</tr>
<tr>
<td>HFZ</td>
<td>62.2%</td>
<td>66.5%</td>
<td>64.3%</td>
<td>60.7%</td>
<td>57.5%</td>
<td>58.9%</td>
</tr>
<tr>
<td>NI - Low Risk</td>
<td>21.4%</td>
<td>20.8%</td>
<td>21.1%</td>
<td>19.9%</td>
<td>26.1%</td>
<td>23.4%</td>
</tr>
<tr>
<td>NI - High Risk</td>
<td>15.9%</td>
<td>11.2%</td>
<td>13.6%</td>
<td>19.0%</td>
<td>15.0%</td>
<td>16.7%</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very Lean</td>
<td>3.0%</td>
<td>2.0%</td>
<td>2.5%</td>
<td>0.9%</td>
<td>1.4%</td>
<td>1.2%</td>
</tr>
<tr>
<td>HFZ</td>
<td>53.2%</td>
<td>56.9%</td>
<td>55.0%</td>
<td>52.1%</td>
<td>57.9%</td>
<td>55.4%</td>
</tr>
<tr>
<td>NI - Low Risk</td>
<td>13.9%</td>
<td>14.7%</td>
<td>14.3%</td>
<td>14.2%</td>
<td>15.7%</td>
<td>15.1%</td>
</tr>
<tr>
<td>NI - High Risk</td>
<td>39.9%</td>
<td>26.4%</td>
<td>28.1%</td>
<td>32.7%</td>
<td>25.0%</td>
<td>28.3%</td>
</tr>
</tbody>
</table>

HFZ = Health Fitness Zone

NI = Needs Improvement Zone
Table 7 Body Composition: Overall classification agreement for measures of body composition (BIA and Skinfold, Skinfold and BMI, BMI and BIA).

<table>
<thead>
<tr>
<th></th>
<th>Very Lean</th>
<th>HFZ</th>
<th>NI - Low Risk</th>
<th>NI - High Risk</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Skinfold</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIA</td>
<td>Very Lean</td>
<td>0.3%</td>
<td>1.3%</td>
<td>0.1%</td>
<td>0.0%</td>
</tr>
<tr>
<td></td>
<td>HFZ</td>
<td>0.7%</td>
<td>51.7%</td>
<td>5.4%</td>
<td>1.0%</td>
</tr>
<tr>
<td></td>
<td>NI - Low Risk</td>
<td>0.0%</td>
<td>8.2%</td>
<td>15.2%</td>
<td>5.3%</td>
</tr>
<tr>
<td></td>
<td>NI - High Risk</td>
<td>0.0%</td>
<td>0.0%</td>
<td>1.7%</td>
<td>9.0%</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>76.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skinfold</td>
<td>Very Lean</td>
<td>0.1%</td>
<td>0.9%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td></td>
<td>HFZ</td>
<td>1.6%</td>
<td>48.7%</td>
<td>8.2%</td>
<td>2.8%</td>
</tr>
<tr>
<td></td>
<td>NI - Low Risk</td>
<td>0.1%</td>
<td>5.1%</td>
<td>5.5%</td>
<td>11.7%</td>
</tr>
<tr>
<td></td>
<td>NI - High Risk</td>
<td>0.0%</td>
<td>0.6%</td>
<td>1.0%</td>
<td>13.7%</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>68.0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BIA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>Very Lean</td>
<td>0.1%</td>
<td>1.2%</td>
<td>0.4%</td>
<td>0.0%</td>
</tr>
<tr>
<td></td>
<td>HFZ</td>
<td>1.6%</td>
<td>49.2%</td>
<td>4.3%</td>
<td>0.2%</td>
</tr>
<tr>
<td></td>
<td>NI - Low Risk</td>
<td>0.1%</td>
<td>6.7%</td>
<td>7.9%</td>
<td>0.0%</td>
</tr>
<tr>
<td></td>
<td>NI - High Risk</td>
<td>0.0%</td>
<td>1.7%</td>
<td>16.1%</td>
<td>10.5%</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>67.7%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HFZ = Health Fitness Zone

NI = Needs Improvement Zone
Table 8 Body Composition: Kappa agreement between measures of body composition (BIA and skinfold, Skinfold and BMI, BMI and BIA) by school type and gender.

<table>
<thead>
<tr>
<th>Kappa of Body Composition</th>
<th>School Type 1</th>
<th>School Type 2</th>
<th>All Together</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>BIA and Skinfold</td>
<td>0.613</td>
<td>0.558</td>
<td>0.630</td>
</tr>
<tr>
<td>Skinfold and BMI</td>
<td>0.445</td>
<td>0.421</td>
<td>0.425</td>
</tr>
<tr>
<td>BMI and BIA</td>
<td>0.351</td>
<td>0.410</td>
<td>0.554</td>
</tr>
</tbody>
</table>
Panel A:

Testing Procedures for the PACER

The PACER (Progressive Aerobic Cardiovascular Endurance Run) is the default aerobic capacity test in FITNESSGRAM. The PACER is a multistage fitness test adapted from the 20-meter shuttle run test published by Ledger and Lamber (1982) and revised in 1988 (Leger et al.). The test is progressive in intensity – it is easy at the beginning and gets harder at the end. The progressive nature of the test provides a built-in warm-up and helps children to pace themselves effectively. The test has also been set to music to create a valid, fun alternative to the customary distance run test for measuring aerobic capacity.

The PACER is recommended for all ages, but its use is strongly recommended for participants in grades K-3. The PACER is recommended for a number of reasons, including the following:

- All students are more likely to have a positive experience in performing the PACER.
- The PACER helps students learn the skill of pacing.
- Students who have a poorer performance will finish first and not be subjected to the embarrassment of being the last person to complete the test.

When you are administering the test to these younger children, the emphasis should be on allowing the children to have a good time while learning how to take this test and how to pace themselves. Allow children to continue to run as long as they wish and as long as they are still enjoying the activity. Typically the test in grades K-3 will only last a few minutes. It is not desirable or necessary to make the children run to exhaustion.

Test Objective
To run as long as possible with continuous movement back and forth across a 20-meter space at a specified pace that gets faster each minute. A 15-meter version of the PACER test has been developed for teachers with smaller sized facilities.

Equipment and Facilities
Administering the PACER requires a flat, non-slippery surface at least 20 meters long, CD or cassette player with adequate volume, CD or audiocassette, measuring tape, marker cones, pencils, and copies of score sheet A or B. Students should wear shoes with nonslip soles. Plan for each student to have a 40- to 60-inch-wide space for running. An outdoor area can be used for this test if you do not have adequate indoor space. There should be a designated area for finished runner and for scorekeepers. You may want to paint lines or draw chalk lines to assist students in running in a straight line.
Test Instructions

- Mark the 20-meter (21-yard, 32-inch) course with marker cones to divide lanes and use a tape or chalk line at the end.
- Make copies of score sheet A or B for each group of students to be tested.
- Before test day, allow students to listen to several minutes of the tape so that they know what to expect. Students should then be allowed at least two practice sessions.
- Allow students to select a partner. Have students who are being tested line up behind the start line.
- The individual PACER CDs have two music versions: one with only the beeps and one with the cadences for the push-up and curl-up tests. Each version of the test will give a 5-second countdown and tell the students when to start.
- Each student being tested should run across the 20-meter distance and touch the line with a foot by the time the beep sounds. The student should take full weight on the foot that is touching the line. At the sound of the beep, the student turns around and runs back to the other end. If some students get to the line before the beep, they must wait for the beep before running the other direction. Students continue in this manner until they fail to reach the line before the beep for the second time.
- A single beep will sound at the end of the time for each lap. A triple beep sounds at the end of each minute. The triple beep serves the same function as the single beep and also alerts the runners that the pace will get faster. Inform students that when the triple beep sounds, they should not stop but should continue the test by turning and running toward the other end of the area.
- Scoring the PACER will require the input of student’s height and weight. Calculation of aerobic capacity requires a score of at least 10 laps (20-meter version).

When to Stop
The first time a student does not reach the line by the beep, the student stops where he or she is and reverses direction immediately, attempting to get back on pace. The test is completed for a student the next time (second time) he or she fails to reach the line by the beep (the two misses do not have to be consecutive; the test is over after two total misses). Students just completing the test should continue to walk and stretch in the designated cool-down area. Note: A student who remains at one end of the testing area through two beeps (does not run to the other end and back) should be scored as having two misses and the test is over.

Scoring
In the PACER test, a lap is one 20-meter distance (from one end to the other). Have the scorer record the lap number (crossing off each lap number) on a PACER score sheet. The recorded score is the total number of laps completed by the student. For ease in administration, it is permissible to count the first miss (not making the line by the beep). It is important to be consistent in the method used for counting with all of the students in the classes.
An alternative scoring method is available. This method does not eliminate students when they miss their second beep (Schiemer, 1996). Using the PACER score sheet B, the teacher establishes two different symbols to be used in recording, such as a star for making the line by the beep and a triangle for not making the line. The scorer then draws a star in the circle when the runner successfully makes the line by the beep and a triangle when the runner fails to make the line by the beep, simply making a record of what occurs. The runners can continue to participate until the leader stops music or until they voluntarily stop running. To determine the score, find the second triangle (or whatever symbol was used). The number associated with the preceding star is the score.

Criterion standards for students in grades K-3 have purposefully not been established. There are concerns regarding the reliability and validity of the test results for very young children. Even with practice, it is difficult to ensure that you children will pace themselves appropriately and give a maximal effort. The object of the test for these younger students is simply to participate and learn about the test protocol.

Students ages 5 to 9 years in grades K-3 do not have to receive a score; they may simply participate in the activity. With the software, you may enter the actual number of laps or enter a score of 0 laps to indicate that they successfully participated in the PACER run. Regardless of the entry, the performance will not be evaluated against a criterion standard. Nine-year-olds in grade 4 may receive a score, and it will be evaluated against a criterion standard. All 10-year-old students receive a score regardless of grade level.

Suggestions for Test Administration

- Both PACER CD’s contain 21 levels (1 level per minute for 21 minutes). During the first minute, the 20-meter version allows 9 seconds to run the distance; the 15-meter version allows 6.75 seconds. The lap time decreases by approximately one-half second at such successive level. Make certain that students have practiced and understand that the speed will increase each minute.
- A single beep indicates the end of a lap (one 20-meter distance). The students run from one end to the other between each beep. Caution students not to begin too fast. The beginning speed is very slow. Nine seconds is allowed for running each 20-meter lap during the first minute.
- Triple beeps at the end of each minute indicate the end of a level and an increase in speed. Students should be alerted that the speed will increase. When students hear the triple beeps they should turn around at the line and immediately continue running. Some students have a tendency to hesitate when they hear the triple beeps.
- A student who cannot reach the line when the beep sounds should be given one more chance to attempt to regain the pace. The second time a student cannot reach the line by the beep, his or her test is completed.
- Groups of students may be tested at one time. Adult volunteers may be asked to help record scores. Students may record scores for each other or for younger students.
- Each runner must be allowed a path 40 to 60 inches wide. It may work best to mark the course.
- If using the audiotape, you may save time by using two tapes and two cassette players. Rewind the first tape while the second group is running the tests, and so forth. Using the CD is a much more efficient method for administering this test item.

(Meredith and Welk, 2010)
Panel B:

Testing Procedures for the One-Mile Run

The one-mile run can be used instead of the PACER to provide an estimate of aerobic capacity ($V_{O_2}^{max}$). For students who enjoy running and are highly motivated, it is a very good alternative assessment. Scoring of the one-mile run will require the input of a student’s height and weight since the calculation of aerobic capacity includes BMI.

Test Objective
To run a mile at the fastest pace possible. If a student cannot run the total distance, walking is permitted.

Equipment and Facilities
A flat running course, stopwatch, pencil, and score sheets are required. The course may be a track or any other measured area. The course may be measured using a tape measure or cross country wheel. Caution: If the track is metric or shorter than 440 yards, adjust the running course (1,609.34 meters = 1 mile; 400 meters = 437.4 yards; 1,760 yards = 1 mile). On a metric track the run should be four laps plus 10 yards.

Test Instructions
Students begin on the signal “Ready, Start.” As they cross the finish line, elapsed time should be called to the participants (or their partners). It is possible to test 15 to 20 students at one time by dividing the group. Have each student select a partner; one is the runner and one is the scorer. While one group runs, partners count laps and record the finish time.

Scoring
The one-mile run is scored in minutes and seconds. You will need to enter a score in the software. Students ages 5 to 9 years in grades K-3 do not have to be timed; they may simply complete the distance and be given a score of 00 minutes and 00 seconds. Regardless of the entry in the software the students ages 5 through 9 years, a performance standard will not be used to evaluate their score. Nine-year-olds in grade 4 should receive a score. All 10-year-olds should receive a score regardless of grade level.

Performance standards for students in grades K-3 have purposefully not been established. There are concerns regarding the reliability and validity of the test results for very young children. Even with practice, it is difficult to ensure that young children will pace themselves appropriately and give a maximal effort. The object of the test for these younger students is simply to complete the 1-mile distance at a comfortable pace and to practice pacing. Remember, the height and weight for each student must be entered in addition to the performance time on the one-mile run. Calculation of aerobic capacity requires a score less than 13:01.
Suggestions for Test Administration

- Call out times as the runners pass the start/stop line to assist students in pacing themselves.
- Preparation for the test should include instruction about pacing and practice in pacing. Without instruction, students usually run too fast early in the test and then are forced to walk in the later stages.
- Results are generally better if the student can maintain a constant pace during most of the test.
- Walking is definitely permitted. Although the objective is to cover the distance in the best possible time, students who must walk should not be made to feel inferior. Encourage students who walk to move at a fast pace, rather than stroll. Attainment of the Health Fitness Zone is the most important factor.
- Have students set a goal before running.
- Students should always warm up before taking the test. It is also important that students cool down by continuing to walk for several minutes after completing the distance. A good suggestion is to have those who have completed the distance do an easy activity (like juggling, hula hoop) while waiting for others to complete the distance. This keeps everyone moving and busy and takes the focus off the slower students who will complete the distance last.
- Administration of the test under conditions of unusually high temperature or humidity or when the wind is strong should be avoided, as these elements may be unsafe or may lead to an invalid estimate of aerobic capacity.
- Counting laps completed and accurately recording the run time can be a problem when a relatively small course is utilized with younger children. Many techniques are acceptable. Pair the students and have the resting partner count laps and record time for the runner. Older students or parents may be asked to assist in recording results for younger students.

(Meredith and Welk, 2010)
Panel C:

**Testing Procedures for Skinfold Measurements**

This section provides information on measuring skinfolds, including suggestions on how best to learn to do skinfold measurements.

**Test Objective**
To measure the triceps and calf (and abdominal for college students) skinfold thicknesses for calculating percent body fat.

**Equipment and Facilities**
A skinfold caliper is necessary to perform this measurement. The cost of calipers ranges from $5 to $200. Both the expensive and inexpensive calipers have been shown to be effective for use by teachers who have had sufficient training and practice.

**Testing Procedures**
The triceps and calf skinfolds have been chosen for FITNESSGRAM because they are easily measured and highly correlated with total body fatness. The caliper measures a double layer of subcutaneous fat and skin.

**Measurement Locations**
The triceps skinfold is measured on the back of the right arm over the triceps muscle, midway between the elbow and the acromion process of the scapula. Using a piece of string to find the midpoint is a good suggestion. The skinfold site should be vertical. Pinching the fold slightly above the midpoint will ensure that the fold is measured right on the midpoint.

The calf skinfold is measured on the inside of the right leg at the level of maximal calf girth. The right foot is placed flat on an elevated surface with the knee flexed at a 90° angle. The vertical skinfold should be grasped just above the level of maximal girth and the measurement made below the grasp.

(Meredith and Welk, 2010)
Panel D:

Testing Procedures for Portable Bioelectric Impedance Analyzers

A number of portable bioelectric impedance analyzer (BIA) devices are now commercially available at a price that is reasonable for most physical education programs (<$100). These devices estimate body composition by measuring the body’s resistance to current flow. A body with more muscle will also have more total body water (and therefore have low resistance to current flow). A body with more fat will have less total body water and greater resistance to current flow. One type of device requires participants to stand on an instrument resembling a bathroom scale while barefoot. Another type of device uses a handgrip system that has participants squeeze handles while extending the arms.

Preliminary results with these devices suggest that they provide similar classification accuracy and body composition estimates as skinfold calipers. Because these devices can produce estimates of body composition faster than a skinfold test and do not require specific skill or experience it may provide a useful alternative to skinfold testing in some schools. The procedure is also less invasive than skinfold testing and may be better accepted in some districts that have specific policies against the use of skinfold calipers.

(Meredith and Welk, 2010)
Panel E:

Testing Procedures for Body Composition

Measurements Protocols
Measure and record the following: standing height (cm) and body mass (kg)

Standing Height – Use the stretch stature method. Stature is the maximum distance from the floor to the vertex of the head. The vertex is defined as the highest point on the skull when the head is held in the Frankfort plane. This position is when the imaginary line joining the orbitale to the tragion is perpendicular or at a right angle to the long axis of the body. Subject is measured with shoes removed.

1. Ask subject to stand with back, buttocks and heels against a stadiometer. Subjects feet should be together and flat on the floor.
2. Place subjects head in the Frankfort plane. Place your hands far enough along the line of the subjects jaw to ensure that upward pressure is transferred through the mastoid processes.
3. Instruct subject to take and hold a deep breath. While keeping the head in the Frankfort plane apply gentle upward lift through the mastoid processes. At the same time place the headboard firmly down on the vertex, crushing the hair as much as possible. Ensure that the feet do not come off the ground and that the position of the head is maintained in the Frankfort plane.
4. Record measurement at the end of the subjects deep inward breath – record stature to the nearest 0.1 cm.
5. Ask subject to step away from the stadiometer.
6. Repeat steps 1 to 4 (For every 10th subject to establish intra-observer reliability).
7. If the 2 measurements differ by more than 0.5 cm then repeat steps 1 to 4.
8. If two measurement record the average value. If three measurements record the median value.

Body Mass – Weigh subject with minimal clothing and with shoes removed.
1. Check the scale is reading zero.
2. Ask subject to stand on the center of scale, without support and with their weight distributed evenly on both feet – record body mass to nearest 0.1 kg.
3. Ask subject to step off the scale.
4. Repeat steps 1 to 3 (For every 10th subject to establish intra-observer reliability).
5. If the 2 measurements differ by more than 0.4 kg then repeat steps 1 to 3.
6. If two measurement record the average value. If three measurements record the median value.