CARL CORCORAN

ST05002150

SCHOOL OF SPORT

UNIVERSITY OF WALES

INSTITUTE, CARDIFF

PREDICTION OF $\dot{V}O_{2max}$ IN RUGBY PLAYERS

Table of Contents

List of Figures

List of Tables

Acknowledgements

The author is forever indebted to the following people for their advice, guidance and general contributions to the study; as well as the motivation and support also given to me throughout this stressful period: Dr Stephen-Mark Cooper (University of Wales Institute, Cardiff), Michele Duford, Jessica Taylor Mark Duford, Lewis Corcoran, Rebecca Rees, Gerald O'Neill, Jack Dainty, Mark Molinaro, Jonny Williams, Laura Williams, Caroline Bickerstaff, Dave Gee, Al Henning, Jason Ousten, Ed Jones, Bill Brant, Michael Jenkin, Audrey Mills, Scott Choucino and Carol Taylor.

This study is dedicated to my Granddad, Peter Mills. He always said I'd be the first one of the family to attend university.

Abstract

This investigation aimed to: (i) assess the reliability of scores from an adapted 20 metre multistage fitness test (MSFT) performed on a dry turf rugby pitch, in which subjects wore studded boots, (ii) develop a calibration model using least squares linear regression which could predict $\dot{V}O_{2max}$ scores (ml kg^{-1} min^{-1}) using scores from the adapted MSFT (cumulative shuttles) and (iii) cross-validate the calibration model by quantifying the agreement between directly measured and predicted $\dot{V}O_{2max}$ scores. A sample of 16 rugby players (12 males and 4 females) took part in the investigation (mean (SD), age 20.8 (±1.2) years, stature 1.81 (±0.08) m and body mass 76.6 (±13.4) kg). Each of the 16 subjects performed two MSFTs in order to establish the reliability of the adapted MSFT (initial test 67 (±21) shuttles, retest 68 (±19) shuttles). The agreement between test and retest scores was quantified using the 95% limits of agreement (LoA) method (bias = -3 (±2) shuttles, t = -0.66, P = 0.517; heteroscedasticity coefficient r = 0.012, P = 0.964; LoA = ±3 shuttles). Eleven subjects were then randomly selected from the group to form a calibration sample. An equation was then formed using average scores from the two MSFTs performed in the reliability stage (66 (±22) shuttles) and $\dot{V}O_{2max}$ scores directly measured using cycle ergometry (44.2 (±8.5) ml kg^{-1} min^{-1}). The equation formed was: $\dot{V}O_{2max}$ (ml kg^{-1} min^{-1}) = 22.3 + (0.327 x MSFT score (shuttles)). The standard error of estimate for the model was calculated as: s_{YX} = ±4.9 ml kg^{-1} min^{-1}.

The remaining five subjects not used in the calibration stage were used to cross-validate the calibration model. MSFT scores, obtained by averaging the two MSFT scores achieved by each subject in the reliability stage (70 (±18) shuttles), were entered into the calibration model and predicted $\dot{V}O_{2max}$ scores were calculated (45.3 (±5.7) ml kg^{-1} min^{-1}). These predicted scores were then compared to $\dot{V}O_{2max}$ scores directly measured using cycle ergometry (44.3 (±7.1) ml kg^{-1} min^{-1}). Their agreement quantified, again using the 95% LoA method (bias = -3.7 (±1.8) ml kg^{-1} min^{-1}, t = -0.99, P = 0.380; heteroscedasticity coefficient r = 0.077, P = 0.902; LoA = ±3.5 ml kg^{-1} min^{-1}). The findings of this study led to the conclusion that an MSFT performed on a dry turf rugby pitch whilst wearing studded boots provides repeatable scores within the sample group of rugby players used in the study. Scores from the adapted MSFT can predict $\dot{V}O_{2max}$ scores using the devised calibration model and the standard error of estimate is lower when compared to the error in the original MSFT study by Léger and Lambert (1982). Despite this, the agreement between predicted and directly measured $\dot{V}O_{2max}$ scores suggested that the calibration model predicted valid $\dot{V}O_{2max}$ scores from adapted MSFT scores for the sample population. Because of the simplicity of the adapted MSFT and its ability to test multiple subjects simultaneously; it was decided that it is a useful test for assessing aerobic fitness of rugby players.

Chapter One - Introduction

1.1 Aerobic requirements of rugby union

Rugby Union is a team game that places large physical demands on players throughout its eighty minute duration. Although the game can be seen to require the ability to perform multiple, intermittent, anaerobic activities; large stresses are predominantly placed upon the aerobic capabilities of the players. Aerobic fitness will account for overall performance throughout a game because of its long duration. Shephard (1994) suggests that activities lasting over one hour, such as a rugby match, are predominantly aerobic activities. It is therefore important, in order to be successful in rugby union, for the coach to be able to assess the aerobic fitness of his players in order to prescribe appropriate training interventions to improve (Luger and Pook, 2004). Shephard (1994) states that maximal oxygen uptake ($\dot{V}O_{2max}$), also known as aerobic power, is the traditional index of aerobic fitness. Aerobic power is a vital component of aerobic function as it defines a subject's upmost tolerance of aerobic exercise (Maud and Foster, 2006). All endurance activities are completed whilst working at a certain percentage of $\dot{V}O_{2max}$; ergo the level of endurance performance can be limited by a low $\dot{V}O_{2max}$ (Wilmore and Costill, 2004). To allow a rugby coach to compare scores among a team fairly, it is likely that the $\dot{V}O_{2max}$ scores for each player will be scaled relative to their individual body mass.

1.2 Measuring $\dot{V}O_{2max}$

The British Association of Sport and Exercise Sciences (BASES; 2007) define $\dot{V}O_{2max}$ as 'the global outcome of the rate at which the body can extract oxygen from the atmosphere via the cardio-pulmonary system, transport it via the cardiovascular system and use it in the skeletal muscle.' The point at which a subject achieves their $\dot{V}O_{2max}$ is defined by the following criteria suggested by BASES: (i) a plateau in oxygen uptake (less than 150 ml min^{-1} oxygen intake in response to increased intensity), (ii) a respiratory exchange ratio above 1.15, (iii) a heart rate close to subject's calculated heart rate max (220-age), (iv) an age-appropriate increase in blood lactate (8-11 mMol L^{-1}) and (v) a rate of perceived exertion above 18 on Borg's scale (The British Association of Sport and Exercise Sciences, 2007).

Direct determination of $\dot{V}O_{2max}$ is the most accurate way of assessing aerobic power; however its use is sometimes considered impractical. This is often because accurate testing requires expensive equipment, laboratory time and trained personnel (Ramsbottom *et al*., 1988). In addition to these factors, the practicality of measuring $\dot{V}O_{2max}$ is also affected by the fact that subjects often have to be tested individually. With rugby union being an example of a team game; it is often multiple subjects (entire teams) that require testing. For analyses of results for multiple players to be most accurate, they must be tested within a small time frame, preferably simultaneously. Due to the limitations of direct $\dot{V}O_{2max}$ testing listed previously, this is often not possible or practical.

1.3 Field Testing

With $\dot{V}O_{2max}$ scores widely accepted as good indicators of aerobic fitness, more practical methods of attaining reliable values were required. Much research has therefore been documented on the use of simple field tests to predict $\dot{V}O_{2max}$. Calibration modelling is used to predict a criterion test score ($\dot{V}O_{2max}$) from the result of a predictor (field) test. Predictions are calculated through the utilisation of linear relationships between the field test scores and scores obtained from a laboratory-determined $\dot{V}O_{2max}$ test. Research has revealed that the use of field tests, such as the multi-stage fitness test, to predict $\dot{V}O_{2max}$, can be highly reliable and valid (Ramsbottom *et al*., 1988; Paliczka *et al*., 1987; Léger and Lambert, 1982), although some studies have predicted $\dot{V}O_{2max}$ scores significantly lower than the corresponding directly measured values (Cooper *et al*., 2005; St. Clair-Gibson *et al*., 1998). The benefits of field testing are vast, with the ability to perform mass tests on multiple subjects with relative ease. There are no requirements for specialist equipment or trained personnel and the tests are much cheaper and are not restricted to a laboratory environment.

Commercially available field tests claim to be able to accurately predict maximal oxygen uptake values. Perhaps the most well known of these being the twenty metre multi-stage fitness test, popularised by Brewer *et al*., (1988).

Many coaches, of many sports including rugby union, believe that the $\dot{V}O_{2max}$ scores, predicted by cross-referencing scores from the multi-stage fitness test with the booklet provided; give an accurate representation of a subject's aerobic fitness and consequently use them as measures of aerobic fitness in their players (Hazeldine and McNab, 1991; Australian Sports Commission, 2000).

A factor that many coaches fail to consider when administering the test is the recreation of the original study's test parameters. Often, tests are completed on grass pitches, concrete or Astroturf surfaces; which all have varying performance implications when compared to the original wooden gymnasium flooring used Brewer *et al*. (1988). In addition to this, tests are also frequently performed on specific populations, like rugby teams. This again has impacts when making predictions based on the data gathered in the original study, which used a more general population. Because Brewer *et al.* (1988) created a calibration model in order to predict $\dot{V}O_{2max}$ scores, the most accurate predictions can only be applied to the sample population and specific test protocol used. Applying the same calibration model to a different group of subjects or test protocol increases the amount of error that can be expected in prediction (Bryman and Cramer, 1996). When a rugby coach administers a MSFT it is often on the grass pitch that most training is performed on. In this case, players are usually wearing studded boots. Footwear is yet another performance influencing factor that previous studies have failed to address.

1.4 Aims of this investigation

To ensure that accurate $\dot{V}O_{2max}$ scores can be predicted for a sample population of rugby players, adjustments to the original MSFT protocol must be made. The modifications include the performance of the MSFT on a grass rugby pitch and the use of appropriate studded footwear; as players would experience in a rugby match. This study therefore aims to (i) determine the test-retest repeatability of the modified MSFT on rugby players, (ii) create a new calibration model to predict $\dot{V}O_{2max}$ scores from MSFT results and (iii) cross-validate the model with a sample from the same population.

Chapter Two – A review of literature

2.1 Introduction

Relevant literature to this investigation has been examined and critiqued in order to highlight gaps in the research that this investigation can aim to fill. This review considers possible limitations of the studies mentioned, as well as identifying correct methods that can and should be employed in this investigation. The analysis of the literature has been split into several sections. This study aims to identify the reliability and validity of the twenty metre multi-stage fitness test as a predictor of $\dot{V}O_{2max}$ in rugby players, through the utilisation of calibration modelling and cross-validation. It is therefore important to understand how calibration models are used, particularly in sporting scenarios. Positive and negative implications of using calibration models must also be highlighted. Literature relating to laboratory testing of $\dot{V}O_{2max}$, as well as substitute field testing will also be looked at. More specifically, research relating to the use of the MSFT as a predictor of $\dot{V}O_{2max}$ will be thoroughly explored. Subsequently, issues with subjects used in testing and methodical approaches used will also be addressed.

2.2 Field testing

Field testing can be defined as the conducting of tests that can be completed outside of a laboratory environment without the requirement of complex equipment or qualified personnel, for data collection or recording (Maud and Foster, 2006).

The use of many laboratory-based fitness tests are considered impractical due to the limited number of subjects able to be simultaneously tested, the requirement of sophisticated, expensive equipment and the use of complex equations to establish scores, as well as the need for qualified personnel and extensive lab time. It has therefore become necessary to develop simplified field tests that are cheap and easy to perform on many subjects, whilst giving an appropriately accurate prediction of what could be achieved in the lab (Baker and Davies, 2002). Many studies have used complex statistical analysis in order to make predictions of expected lab test scores, based upon results from field tests. Simplified field testing can be substituted for many lab-based fitness tests using these methods; examples include tests for aerobic power, anaerobic power, muscle power and even body densities. The field tests used vary between studies, as does the criterion that is hoped that it will predict. For example, Baker *et al*. (1993) attempted to quantify the ability to use scores from a 40 metre shuttle run to predict anaerobic performance. They found moderate correlations of r = -0.67 and r = -0.75 between the shuttle run scores and scores from two standard lab tests for anaerobic power and capacity. They concluded that scores from the 40m shuttle run could be used as a predictor of anaerobic performance.

Results are not always highly correlated between scores from field and laboratory tests. Tharp *et al*. (1985) attempted to prove a correlation between sprint / run times (seconds) and scores from the Wingate anaerobic test (work or kgm).

Results showed only a moderate correlation between sprint times (seconds) and Wingate scores (kgm) of r = -0.53; and an even weaker correlation between run times (seconds) and Wingate scores (kgm) of r = -0.28. This may suggest that the scores from the field tests chosen may not have been appropriate for predicting the desired criterion scores.

It is not always immediately obvious which field tests will give scores that correlate highest with the desired lab test scores. Studies such as that by Aziz *et al*. (2003) compared scores from more than one field test with scores from a lab test, to see which one had the better relationship. Their study compared the use of the MSFT and the yo-yo intermittent endurance test (YIET) with a treadmill $\dot{V}O_{2max}$ test. They aimed to discover which field test was most appropriate for predicting the $\dot{V}O_{2max}$ of soccer players. It was proven that the MSFT was highly correlated with directly determined $\dot{V}O_{2max}$, whilst the YIET was not. Because soccer is a team game in the football code, it is comparable to rugby union. It is therefore assumed that the MSFT would be a good measure of $\dot{V}O_{2max}$ in rugby players and this study aims to assess that assumption.

The 20m MSFT was introduced by Léger and Lambert (1982) as a possible predictor of $\dot{V}O_{2max}$, but had a large standard error of estimate (s_{YX} = ±5.4 ml kg^{-1} min^{-1}). Kline *et al*. (1987) suggested an alternative field test to predict $\dot{V}O_{2max}$. A calibration model was created to estimate $\dot{V}O_{2max}$ from a one mile track walk, gender, age and weight.

They still experienced a relatively high standard error of estimate (±4.5 ml kg^{-1} min^{-1}) when compared to Ramsbottom *et al*.'s (1988) modification and re-appraisal of the MSFT (s_{YX} = 3.5 ml kg^{-1} min^{-1}).

2.3 $\dot{V}O_{2max}$ in predicting performance

$\dot{V}O_{2max}$ values are used widely as indicators of aerobic fitness. High levels of aerobic fitness can be beneficial to successful performance in many sports, including rugby union (Scott *et al*., 2003). Many studies have explored the relationship between directly measured $\dot{V}O_{2max}$ scores and sports performance and have found that $\dot{V}O_{2max}$ does link to some aspects of sport performance. There have been numerous different protocols used for testing and studies have also been geared toward a variety of specific sports. Notably, it is easier to establish relationships between $\dot{V}O_{2max}$ and sport performances that are easily quantified. For instance, it is easier to quantify performance of an individual running a 10km race (time), than an individual's performance in a rugby match, where such obvious measures of performance do not exist.

However, studies such as that by Rampinini *et al*. (2006) used a video-computerised, semi-automatic match analysis recognition system to quantify performance in a game of soccer and compared scores to those obtained on simple field tests. When this complex equipment is not available; subjective performance quantification is often employed.

Gabbett *et al*. (2007) attempted to link testable performance indicators, including $\dot{V}O_{2max}$, with rugby league performance. Results found no significant difference between $\dot{V}O_{2max}$ scores of players of varying ability ($P > 0.05$). The obvious limitation to the methods involved in the study, was that performance or playing ability, was graded rather subjectively by a coach. Conversely, a study by Van Schuylenbergh *et al*. (2004) was able to use $\dot{V}O_{2max}$ data and selected other physiological criteria to accurately predict times in a sprint triathlon race, eliminating the subjective element. Palickza *et al*. (1987) found a strong correlation between $\dot{V}O_{2max}$ and 10km race time (r = -0.95) and concluded that $\dot{V}O_{2max}$ scores were valid indicators of performance.

2.4 Calibration models

Calibration models use statistical methods of least squares linear regression to form equations, from which scores from one test can be used to predict scores in another test. In sport sciences, calibration models are most often used to predict scores for complex laboratory tests, from scores obtained on simpler field tests or scores determined by a performance. Calibration models are used in many other instances (e.g. medicine or engineering), not only to increase the practicality of testing, but also to limit health risk factors of criterion testing by utilisation of a good predictor test (Valderrama *et al*., 2007; Nijsten *et al*., 2007; Tomczak *et al*., 2007). The main advantage of the use of calibration modelling is that it allows output scores to be given in different units to the input scores.

This makes it easier to substitute the simpler field tests (perhaps measured in time or shuttles) for more complex laboratory tests (perhaps measured in watts or L min^{-1}), based on calculated predictions. The obvious disadvantage therefore, is that the new scores are only predictions and are therefore subject to error; known as the standard error of estimate (s_{YX}).

2.5 The specific use of the multistage fitness test as a predictor of $\dot{V}O_{2max}$

The MSFT has emerged as one of the most commonly used field tests implemented to estimate aerobic fitness. This is not the first study to use the MSFT as a predictor of $\dot{V}O_{2max}$ scores however; or assess the reliability and validity of these tests. Many studies over the last 20 years have aimed to confirm or disprove the findings of Léger and Lambert (1982); who were the first to use the MSFT to predict $\dot{V}O_{2max}$ (Lamb and Rogers, 2007; Cooper *et al*., 2005; Stickland *et al*., 2003; St. Clair-Gibson *et al*., 1998; Léger and Gadoury, 1989; Ramsbottom *et al*., 1988; Léger *et al*., 1988; Palickza *et al*., 1987). Results have varied among these studies, but most suggest that the MSFT is a valid and reliable predictor of $\dot{V}O_{2max}$. The differences in results, however, might be attributed to several factors that vary between studies; these include: the number of subjects tested, their age, gender and levels of fitness, the use of specific populations, data analysis methods and methods of determining $\dot{V}O_{2max}$.

The original study by Léger and Lambert (1982), used 91 subjects in total and the sample was heterogeneous (32 females and 59 men). Other studies using heterogeneous samples had sample sizes as high as 122 (Stickland *et al*., 2003) and as low as 35 (Lamb and Rogers, 2007). Sample sizes should be as high as possible in order to limit the occurrence of results due to random chance. Sample sizes can be limited, however, by the amount of volunteers gathered, the time frame in which the study takes place and the practicalities of testing large numbers of subjects.

The inclusion of male and female subjects in studies like these has a large influence on the correlation between MSFT scores and $\dot{V}O_{2max}$ scores. The correlation is likely to be stronger when using heterogeneous groups, rather than homogeneous groups, because the inclusion of both genders widens the range of characteristics of the subject group, increasing the range of the covariates that might have an impact on the results of the tests (body mass, stature, fitness *etc*.). This is signified by the fact that studies using heterogeneous samples experienced correlations between MSFT scores and $\dot{V}O_{2max}$ values, as high as r = 0.92 (Ramsbottom *et al*., 1988), whilst studies that used homogeneous samples returned correlations as low as r = 0.67 (St. Clair-Gibson *et al*., 1998).

Léger and Lambert (1982) gave no indication as to the level of physical activity or participation in specific sports, for the sample and a wide range in age was experienced with a standard deviation of ±7.4 years from the arithmetic mean age of 26.1 years.

It has been proven by Léger *et al*. (1988) that age influences MSFT – $\dot{V}O_{2max}$ correlation coefficients. In adults aged 18-50 years, the correlation was much higher (r = 0.90) than in children aged 8-19 years (r = 0.71). It is therefore important to limit the range of ages within the sample group, in order to decrease the influence that age has on the correlation. Fitness levels of subjects can also affect generalised prediction equations. In the study by Lèger and Lambert (1982), all of the subjects were healthy but not necessarily highly trained. This increased the possibility of a wide range in scores across the sample group. This strengthened the correlations. Studies that have used sport specific or well trained subjects have experienced weaker correlations. St. Clair-Gibson *et al*. (1998) tested squash players and runners and found MSFT – $\dot{V}O_{2max}$ correlations of only 0.61 and 0.71 respectively. This is due to the fact that most of the subjects had similar fitness levels and training adaptations. This decreased the range of scores experienced across the group and effectively weakened the correlations. It is this issue that has led to drastic under predictions of $\dot{V}O_{2max}$ scores by previous studies (Léger and Lambert, 1982 and Ramsbottom *et al*., 1988). This is because many athletes, fitter than the original subjects in the studies, have a $\dot{V}O_{2max}$ beyond the range of values attained by those original subjects (St. Clair-Gibson *et al*., 1998).

Conversely, a study by Siegler *et al*. (2006) investigated the effects of applying testing protocols developed on elite soccer players, to non-elite players. No significant relationships were found between elite and non-elite fitness test scores.

Some sports, including rugby union, have a wide range of fitness requirements for successful performance. In the instance of rugby union, physiological requirements vary between positions. Scott *et al*. (2003) reported that backline players had, on average, $\dot{V}O_{2max}$ scores 7.1 ml kg^{-1} min^{-1} higher than forward players. Again, this can cause a wide range in recorded $\dot{V}O_{2max}$ scores even within a sport-specific population.

The MSFT protocol used by Léger and Lambert (1982) started participants at a running speed of 8 km h^{-1} and used incremental stages with durations of two minutes; with the speed increasing by 0.5 km h^{-1} at the end of each stage. Subsequent studies have proven the validity of an altered test protocol (Léger *et al*., 1988; Ramsbottom *et al*., 1988), in which subjects begin at a running speed of 8.5 km h^{-1} with 0.5 km h^{-1} increments every minute, not two. This decrease in stage time was introduced in an attempt to limit subjects losing motivation or becoming bored, causing subjects to prematurely withdraw before reaching their maximal physiological exertion (Léger and Gadoury, 1989). The surface on which the test is conducted and the footwear used by the subjects are protocol variables rarely considered in studies using the MSFT.

Whilst some investigations have specified sprung wooden gymnasium flooring as the test surface (Cooper *et al*., 2005; Ramsbottom *et al*., 1988, Léger and Lambert, 1982); footwear used is consistently overlooked. These factors can have significant influences on results.

For example, a test conducted on grass whilst wearing studded boots could have very different outcomes to a test performed in a gymnasium whilst wearing good running shoes. The obvious effects that footwear can have on the performance of the MSFT have yet to be investigated or even considered in test protocols.

The method by which $\dot{V}O_{2max}$ is determined also varies amongst studies. Earlier studies used a method detailed by Léger *et al*. (1980) called retroextrapolation (Léger *et al*. 1988; Léger and Lambert, 1982). In this method, four $\dot{V}O_2$ values were taken every 20 seconds during recovery after completing the MSFT. Exponential least squares regression of these values was then used to calculate $\dot{V}O_2$ at time zero of recovery. Assuming that the test was completed maximally, this extrapolated value could be considered as the subject's $\dot{V}O_{2max}$. This method uses regression however, and is therefore subject to error (s_{YX} = ±5.4 ml kg^{-1} min^{-1}).

Alternative methods of $\dot{V}O_{2max}$ determination have been suggested in studies by Uth *et al*. (2004) and Vehrs *et al*. (2007) and have favourable error when compared to the retroextrapolation method. Uth *et al*. (2004) suggested that $\dot{V}O_{2max}$ could be estimated relatively accurately (s_{YX} = ±2.7 ml kg^{-1} min^{-1}) from the ratio between heart rate max and resting heart rate, in well-trained men. This method has not been validated for application to other groups however.

Vehrs *et al*. (2007) detailed a method of determining $\dot{V}O_{2max}$ in fit adults through use of a single stage sub-maximal treadmill run. A multiple linear regression model was then applied to scores and used to predict $\dot{V}O_{2max}$. The error experienced when this method was employed was even smaller than that experienced in previous studies (s_{YX} = ±2.52 ml kg^{-1} min^{-1}).

Direct measurements of $\dot{V}O_{2max}$ values via laboratory tests are more accurate and thus, many studies have since favoured the use of maximal, progressive treadmill run test protocols for $\dot{V}O_{2max}$ determination (Palickza *et al*., 1987; Ramsbottom *et al*., 1988; St. Clair-Gibson *et al*., 1998 and Stickland *et al*., 2003). No studies have tested the agreement between scores on the MSFT and $\dot{V}O_{2max}$ scores directly determined from cycle ergometry. It has been reported by Cooper *et al*. (2005) and Grant *et al*. (1995), that previous calculations by Ramsbottom *et al*. (1988) and Léger and Lambert (1982) systematically under predict $\dot{V}O_{2max}$. As stated by Basset and Boulay (2000), $\dot{V}O_{2max}$ values directly obtained using cycle ergometry are up to 6.8% lower than when treadmill protocols are used. It is therefore possible that the use of cycle ergometry to determine $\dot{V}O_{2max}$ will agree more favourably with MSFT scores than treadmill protocols. In addition, research by Hill *et al*. (2003) and Carter *et al*. (2000) found that oxygen uptake kinetics were similar between cycle ergometry and treadmill run $\dot{V}O_{2max}$ protocols, with the only difference being a significantly higher $\dot{V}O_2$ slow component in the cycle ergometry protocol.

This was explained by higher recruitment of type II muscle fibres in the end phase of the cycling than the running, thus indicating a higher maximal exertion when cycle ergometry was used.

2.6 Reliability, validity and methods of their quantification

Cross-validation is another important consideration that is often overlooked in the research. Calibration models can only give their most accurate predictions to the same group that was used to create them. Application of a calibration model to subjects outside of this group results in larger errors in prediction or shrinkage (Bryman and Cramer, 1996). Cross-validation of the model involves applying it to other subjects outside of the calibration group. Scores from the predictor test are used to give predicted scores for the criterion test. The predicted score is compared with the actual measured score and their agreement quantified (Cooper *et al*., 2004). Cross-validation gives an indication as to how well the model can predict criterion test scores from predictor test scores. Many studies involving the MSFT and $\dot{V}O_{2max}$ did not cross-validate their prediction models (Léger and Lambert, 1982; Ramsbottom *et al*., 1988). Cooper *et al*. (2004) used a modified version of the MSFT to predict anaerobic capacity in netball players. A calibration model was created in this study and it was cross-validated. It was concluded that, because the agreement between predicted and measured scores was sufficient, the adapted MSFT could be used as a valid predictor of mean power output.

It has been highlighted by Cooper *et al*. (2005) that some statistical analyses used in previous studies to quantify the reliability and validity of the MSFT as a predictor of $\dot{V}O_{2max}$, may not be truly representative of the agreement between tests. The criticised methods include the use of correlation coefficients (Palickza *et al*., 1987) and one-way analysis of variance (ANOVA), utilised by Dabonneville *et al*. (2003). An alternative method of analysis was also suggested and implemented by Cooper *et al*. (2005), as the 95% limits of agreement method (Bland and Altman, 1986). This method is arguably more appropriate and was also used by Lamb and Rogers (2007) to quantify validity.

2.7 Aims based on literature review

Based on the literature in the area, this study aims to validate the MSFT as a predictor of $\dot{V}O_{2max}$ in male and female rugby players, with participants wearing studded boots on a grass turf surface for the field test. $\dot{V}O_{2max}$ will be determined by a maximal, progressive test using cycle ergometry.

Chapter Three - Methods

3.1 Subjects

The sample group of 16 subjects, used in this investigation, consisted of 12 males and 4 females (mean (SD): age 20.8 (±1.2) years, stature 1.81 (±0.08) m and body mass 76.6 (±13.4) kg), all of whom had played rugby union to at least school first team level. Before participant recruitment began, ethical approval was sought and attained, from the University of Wales Institute, Cardiff (UWIC) Research Ethics Committee (UREC), for the proposed tests involved. Test protocols were then subjected to a risk assessment. Risk factors involved in testing were considered and an overall risk score calculated. A sufficiently low score meant that the test protocols were relatively safe. Before testing, each subject was asked to complete a pre-exercise health questionnaire (Par-Q); in order to assess their physical condition and ability to take part in the investigation. They were then advised of the nature of the tests involved before signing forms of informed consent (Par-Q and informed consent example forms are provided in Appendix A, along with ethics and risk assessment forms). Data collection began by obtaining basic physiological characteristics of each subject, stature and body mass. These were collected by standardized procedures (Lohman, 1988). This was done through use of a Holtain Fixed Stadiometer (Holtain LTD, United Kingdom) and Seca 770 Digital Scales (Vogel and Halke, Germany) respectively. These results were then recorded together with the relevant subject's age.

3.2 Stages of data collection

The aim of this investigation was to validate the use of a predictor test (MSFT) to accurately predict the results of a criterion test ($\dot{V}O_{2max}$ test). The calibration modelling involved had been utilised in a past study by Cooper *et al*. (2004) and this investigation followed the same three part investigation structure. The first stage aimed to indicate the reliability (test-retest repeatability) of the MSFT. Reliability relates to the repeatability or consistency of the test and is an integral part of validity (Thomas *et al*., 2005). A test cannot be considered valid if it is not reliable. In this stage, all sixteen subjects completed two MSFTs within a three week testing period. The MSFT scores (stage:shuttle) were also expressed as the number of cumulative shuttles completed. Cumulative shuttles completed have been used for all analysis involving MSFT scores. Scores from the test and retest were tested for their agreement. This was done using the 95% limits of agreement (LoA) method detailed by Bland and Altman (1986) and explained in more detail in section 3.6.

Stage two gave the basis for the formation of the calibration model. Eleven subjects, randomly selected from the group, performed a $\dot{V}O_{2max}$ test in a laboratory within the three week testing period. An average score of the two MSFTs performed by each subject in stage one, was taken as their MSFT score for use in the second stage. Results collected from the laboratory test were used as criterion data for the calibration model, whilst the average results of the MSFTs were used as predictor data.

The results were analysed to see if a linear relationship between the variables existed before using methods of linear regression to generate the calibration model (Bryman and Cramer, 1996).

The five subjects not used in stage two were used in stage three for the cross-validation of the calibration model. Each of the five subjects performed a laboratory-based $\dot{V}O_{2max}$ test, again within the three week testing period. An average of each of the five subject's two MSFT scores used in stage one, was taken as their MSFT score for stage three. The average results from the MSFTs were then inserted into the calibration model (developed in stage 2) in order to obtain a set of predicted $\dot{V}O_{2max}$ scores. These predicted scores were then tested for their agreement with the lab-determined scores; again the 95% LoA method was used (Bland and Altman, 1986).

3.3 Criterion (laboratory tested $\dot{V}O_{2max}$) test protocol

A strict protocol was followed by all participants throughout the laboratory $\dot{V}O_{2max}$ testing. This was to ensure consistency and reliability of results. The testing had to be consistent in order for the test to be considered valid. Keeping to a strict protocol limits the occurrence of measurement error (Thomas *et al*., 2005). Each subject first sat on the Monark 874 cycle ergometer (Monark Exercise AB, Sweden) to test and adjust the correct height of the seat. This was judged by the knee angle being as close to ninety degrees as possible at the top of a revolution, whilst in a seated position.

Before the test began, each participant was once again made aware of the test protocols. Also at this time, eight 200L Douglas bags (Hans Rudolph Inc., USA) were prepared for collection of expired gas as this was more than any subject was expected to fill. They were emptied using a Harvard dry gas meter (Harvard Apparatus, USA) and the bags were connected to a two-way respiratory valve and mouth piece via 32mm bore respiratory tubing (Hans Rudolph Inc., USA). An unloaded cycle warm-up lasting four minutes was set as the standard warm-up for all participants (Buchfuhrer *et al*., 1983).

After the warm-up, participants, when ready, began cycling. They accelerated to the correct revolutions per minute in accordance with the standard (70 revolutions per minute (rpm) for males and 50rpm for females) and maintained this. Resistance at the start of the first stage was the 1kg weight of the cycle ergometer's load basket. When the correct speed was reached timing began on a Casio stopwatch. The test consisted of consecutive, three minute stages of increasing intensity until the subject was exhausted and unable to continue (Karpman, 1987). The test was continuous with no rests, but subjects did receive verbal motivation as this encouraged maximal exertion (The British Association of Sport and Exercise Sciences, 2007). Resistance was increased by 0.5kg after every third minute, in conjunction with the timekeeper, whilst rpm remained constant. Expired gas was collected for every third minute of exercise. The timekeeper would give a thirty second warning to the subject to allow them to put the mouth piece in and a nose clip on. This was shortly followed by a five second countdown so that the valve on the Douglas bags could be opened accurately on time.

Another five second countdown towards the end of each stage would result in the valve on the Douglas bag being closed. These processes were repeated for every third minute until either the subject reached their age determined heart rate maximum (220-age) or the subject was unable to continue (volitional exhaustion); at either point testing was ceased (The British Association of Sport and Exercise Sciences, 2007). Heart rate was monitored throughout the test through use of a Polar S610 heart rate monitor (Polar Electro Oy, Finland). Resting heart rate before testing and peak heart rate during the test were recorded for each subject (presented in Appendix B). All subjects exceeded 90% of their heart rate maximum ensuring maximal effort during the test.

The concentrations of expired oxygen and carbon dioxide in the Douglas bags were analysed through use of a dual gas analyser (Servomex Model 1440c). The dual gas analyser was calibrated before use by analysing gases of known oxygen and carbon dioxide concentration and adjusting the machine accordingly. One litre of the expired gases was used to obtain the average values for expired oxygen (FEO_2) and carbon dioxide ($FECO_2$) for each stage of the test completed. The total volume of expired gas within each measured stage was obtained using a Harvard dry gas meter. One litre was then added to this value to replace the litre taken by the dual gas analyser. This process was repeated for each stage and the results recorded.

These results were then entered into a spreadsheet that calculated $\dot{V}O_{2max}$ in ml kg^{-1} min^{-1} and L min^{-1}. This was done using the following equations: (i) $\dot{V}O_{2max}$ (L min^{-1}) = (minute ventilation ($\dot{V}_E$, measured in L) x (20.93 x FEO_2)) / 100; (ii) $\dot{V}O_{2max}$ (ml kg^{-1} min^{-1} = ($\dot{V}O_{2max}$ (L min^{-1}) / body mass of subject (kg)) x 1000.

3.4 Scaling

$\dot{V}O_{2max}$ values, within sports science environments, have often been scaled for individuals into millilitres per kilogram of body mass per minute (Shephard, 1994). This method of scaling is subject to criticism and it is suggested by Tanner (1949) that use of ratio standards such as this are misleading; unless 'Tanner's exceptional circumstance' ((cv_X/cv_Y) = r_{XY}) is satisfied. If this equation is met, then measurements can be accurately adjusted for a subject's body size. Despite this fact, it was decided that ml kg^{-1} min^{-1} was the most appropriate units to use because the MSFT cross-reference table, provided with Brewer *et al.*'s (1988) version of the MSFT, expresses $\dot{V}O_{2max}$ in this form. Therefore, for best comparisons between studies, the same units of $\dot{V}O_{2max}$ measurement are preferable.

3.5 Predictor (multi-stage fitness test) test protocol

After performing a standardized warm-up, consisting of a two lap jog around the rugby pitch and a mixture of relevant static and dynamic stretches, subjects completed a slightly modified version of the MSFT. Despite being pioneered as a fitness test for rugby players originally by Léger and Lambert (1982), the test protocol used was essentially identical to the protocol detailed in Brewer *et al*. (1988). The only difference was the surface on which the test was undertaken. To make this investigation more specific to rugby players, the MSFTs were performed on a dry, turf rugby pitch, as opposed to the sprung wooden flooring, suggested by Brewer *et al*. (1988). Participants were also asked to wear studded boots for the test. This not only allowed for a closer simulation of rugby playing conditions, but also aided safety because of the extra grip given by studded boots.

The test area was set up in the centre of a rugby pitch, using the opposing ten metre lines each side of the half way line as the twenty metre shuttle distance. This distance was confirmed by use of a tape measure prior to testing. Subjects were instructed to follow the audio cues from the investigator's laptop in order to judge the speed to run between the lines. TeamBeep (Bitworks engineering, http://www.rugbycoach.com/fitness/test/20msrt.htm), was the software used to emit the audio cues needed for the test. The software was downloaded to a laptop computer and used to administer the MSFT.

This version of the test was chosen over the versions available on audio cassette or compact disk because the timing of the beeps was extremely accurate and the error could be calculated by the software after each test. The downloaded version of the test suffered limited timing error and required no manual calibration. From a test run of the program, the average timing error incurred on each shuttle was ±0.04s. This compared favourably to the study by Cooper *et al*. (2004) where a timing of error of ±0.25s every 30 seconds was accepted using the cassette version of the test.

The version of the test used had twenty levels; lasting one minute each and beginning at a running speed of 8.5 km h^{-1}. This increased by 0.5 km h^{-1} after every stage. Subjects were encouraged to run to volitional exhaustion, completing as many shuttles as they could, whilst continuing to turn in time with the beeps. Multiple subjects performed the test together, adding a motivational edge. Subjects were disqualified from the test if they failed to complete two consecutive shuttles within the audio defined time scales. At the point of voluntary withdrawal or disqualification, the relevant subject's result was recorded as a stage level and shuttle number, only fully completed shuttles were considered.

3.6 Statistical analysis

Statistical analyses in this investigation were computed by the use of the software program Minitab 14. Before parametric analyses of any of the data used in this investigation could be implemented, normality of the residual errors of scores was quantified through use of Anderson-Darling Normality tests. As mentioned previously, the first stage of the investigation aimed to quantify the agreement between results in both the initial MSFT and retest MSFT. First of all, a paired *t*-test was performed on the test and retest scores in order to quantify any bias between them (mean and standard deviation differences). This also tested the null hypothesis of no difference between the means of the tests (Vincent, 1999). A graph was then plotted of the mean of each subject's test and retest scores; against the residual errors between each subject's test and retest scores. This is known as a *Bland-Altman* plot (Bland and Altman, 1986). The use of regression assumed that the dispersion of results was homoscedastic. If the amount of scatter around the line of best fit varied markedly at different points, the data was said to be heteroscedastic (Bryman and Cramer, 1996). Heteroscedasticity was quantified by obtaining a zero order correlation between absolute differences between test-retest scores and the means of the test-retest scores. A correlation close to zero meant that the 95% LoA, were expressed as ±1.96 x SD_{diff} (obtained from the paired *t*-test). Including the bias, the 95% LoA was written as $\bar{x}_{diff}$ ±(1.96 x SD_{diff}). These upper and lower limits were then superimposed on the *Bland-Altman* plot.

A Pearson's correlation coefficient was also calculated to quantify the relationship between test and retest scores. This was mainly for easy comparison to previous investigations that had used correlation coefficients to assess reliability.

The first statistical method involved in developing the calibration model in stage two of the investigation was determining a zero order correlation between predictor test and criterion test scores. This tested the null hypothesis of no linear relationship between the two test scores. The coefficient of determination (CoD, R^2) was then determined by use of the following equation: CoD % = r^2 x 100. The statistical analysis software also gave an adjusted CoD value. This percentage score indicated the amount of common variance between the criterion and predictor test results and was visually presented by use of a Venn diagram (Bryman and Cramer, 1996; Vincent, 1999). If the correlation was high, then the CoD would be high too. A calibration model was formulated using least squares methods of linear regression (Cooper *et al*., 2004). The end equation was therefore a straight line equation (Y = a + bX). Thus, for accurate prediction of a result from a criterion test (Y) using the result of a predictor test (X), there needed to be a strong linear correlation (a value of r = 1 would have meant that Y could have been exactly predicted from X as there was 100% common variance). From the straight line equation, we wanted to predict Y using X.

To do this, the a and b components of the equation were required. The gradient of the line (b) is the change in the value of *Y* when the *X* value is increased by one unit. The equation for b is therefore as follows: b = $r_{XY}(s_Y/s_X)$, with r_{XY} being the zero-order correlation, s_Y being the standard deviation of the *Y* scores and s_X being the standard deviation of the *X* scores. The b value helped calculate the a value (*Y* intercept) using the equation: a = $\bar{x}_Y - (b \times \bar{x}_X)$. $\bar{x}_Y$ denoted the mean of all of the *Y* scores and $\bar{x}_X$ was the mean of the *X* scores. With the a and b components calculated, a result from the predictor test (*X*) inserted into the calibration model would give a predicted value for the criterion test (*Y*). Because it is highly unlikely for the CoD to be 100%, calibration models with less than 100% common variance of tests are subject to error.

This error is the standard deviation of the residual scores and is known as the standard error of estimate (s_{YX}). This is calculated with the equation $s_{YX} = \pm s_Y \sqrt{(1 - r_{XY}^2)}$. This gives a range due to error above and below the predicted score, in which the actual score might lie. To be 95% sure of this, s_{YX} is multiplied by ±1.96 to give $\pm s_{YX}$. As calibration models rely on methods of least squares linear regression (parametric analysis) it is of paramount importance that the residual errors derived from the regression are normally distributed (Altman, 1999). Therefore an Anderson-Darling test of normality was performed to confirm that the scores were normally distributed and that parametric analyses could be used on the data.

The final part of the investigation aimed to cross-validate the calibration model. Scores from the MSFT were substituted into the calibration model (developed in stage two) in order to obtain a predicted $\dot{V}O_{2max}$ score. These predicted scores were then compared to actual scores obtained from their laboratory-based $\dot{V}O_{2max}$ test; and tested for their agreement. Like in stage one, the 95% LoA method was applied. A paired *t*-test was performed and the normality of residuals, as well as the heteroscedasticity coefficient were calculated. The results were then summarised in a *Bland-Altman* plot, fully quantifying the 95% LoA. The relationship between predicted and measured $\dot{V}O_{2max}$ was also quantified by a Pearson's correlation coefficient. This type of validity was identified as concurrent criterion validity (Thomas *et al*., 2005).

Chapter Four - Results

All raw data collected and used for statistical analysis during the investigation can be found in Appendix B. Mean and standard deviation values were also calculated for each measured variable used at each stage of the investigation (see Table 1)

Table 1 - Means (± standard deviations) for each variable calculated and used at each stage of this investigation.

		Stage 1 MSFT reliability *n* = 16	Stage 2 Calibration Model *n* = 11	Stage 3 Cross-validation *n* = 5
Age (yrs)		20.8 (1.2)	20.6 (1.4)	21.2 (0.3)
Stature (m)		1.81 (0.08)	1.82 (0.09)	1.79 (0.07)
Mass (kg)		76.6 (13.4)	77.6 (15.1)	74.2 (9.6)
MSFT	Test	67 (21)		
(shuttles)	Retest	68 (19)		
Average of stage 1 MSFTs			67 (22)	70 (18)
VO_{2max} (ml kg^{-1} min^{-1})			44.2 (8.5)	44.3 (7.1)
Predicted VO_{2max}				45.3 (5.7)

4.1 Stage One

In the first stage of the investigation, two MSFTs were completed by each of the sixteen subjects in order to assess its test-retest repeatability (reliability). Results from the initial test and the retest were 67 shuttles (±21 shuttles) and 68 shuttles (±19 shuttles) respectively. A paired *t*-test was then used to assess the null hypothesis (H_0) of no bias between the means of the test and retest MSFT scores (H_0: $\bar{x}_T = \bar{x}_{RT}$). Non-significant bias was evident: $\bar{x}_{diff}$ (SD_{diff}) = -3 (±2) shuttles; t = -0.66, P = 0.517. An Anderson-Darling normality test found that the residual errors (difference between test and retest scores) were normally distributed (A^2 = 0.30, P = 0.538). The heteroscedasticity coefficient was sufficiently close to zero (r = 0.012, P = 0.964), and thus the limits of agreement were expressed as ±1.96 x SD_{diff} (±3 shuttles).

A Pearson's correlation between test and retest MSFT scores showed a strong relationship (*r* = 0.98; *P* < 0.05). These results are presented in Figure 1.

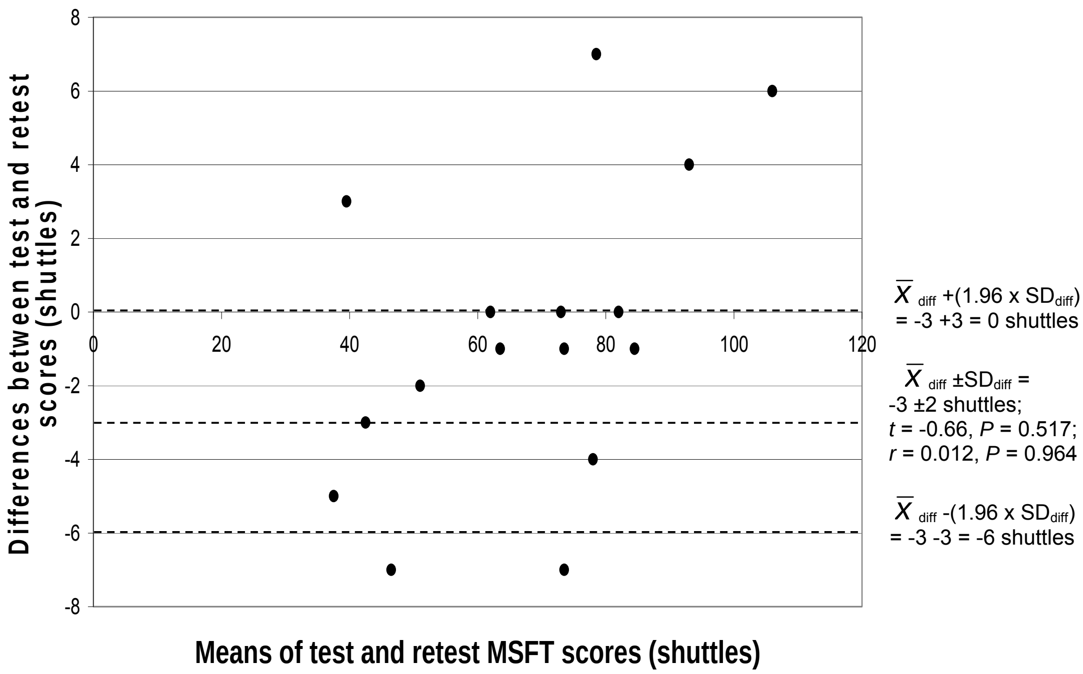

Figure 1 - The reliability of the MSFT was summarised in a Bland-Altman plot. Values for heteroscedasticity and bias are superimposed on the plot.

4.2 Stage Two

Eleven of the sixteen subjects were randomly selected to form a calibration sample. $\dot{V}O_{2max}$ scores determined from the laboratory-based cycle test were then obtained from each subject in the calibration sample. Mean $\dot{V}O_{2max}$ was 44.2 ml kg^{-1} min^{-1} (±8.5 ml kg^{-1} min^{-1}). An average of each subject's two MSFT scores (used in stage one of this investigation) was then calculated and used as their MSFT score in this stage ($\bar{x}$ (SD) = 67 (±22) shuttles).

The null hypothesis of no linear relationship (H_0: $r = 0$), between $\dot{V}O_{2max}$ (ml kg^{-1} min^{-1}) and MSFT (shuttles) scores was tested by obtaining a zero order correlation coefficient. A significant positive correlation was found ($r = 0.835$; $P = 0.001$).

A plot of $\dot{V}O_{2max}$ scores versus MSFT scores confirmed the existence of a linear relationship (seen in figure 3). With a strong linear relationship identified, it was decided that a calibration model could be formed in order to calculate $\dot{V}O_{2max}$ scores in ml kg^{-1} min^{-1} (*Y*; criterion), from MSFT scores in shuttles (*X*; predictor). The coefficient of determination (R^2) was calculated at 69.8% (adjusted to 66.4%) and a Venn diagram was drawn to display the common variance (see Figure 2).

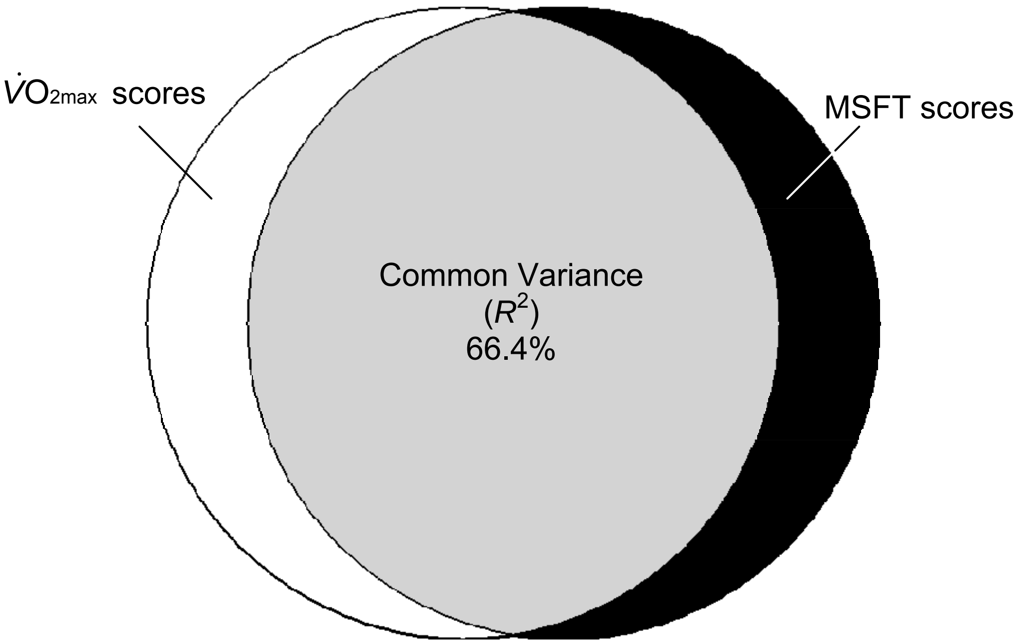

Figure 2 – Common variance (CoD adjusted) between $\dot{V}O_{2max}$ and MSFT scores.

The following equation (calibration model) was then formulated using least squares methods of linear regression in the form *Y* = a + b*X*: $\dot{V}O_{2max}$ (ml kg^{-1} min^{-1}) = 22.3 + (0.327 x MSFT score (shuttles)). The standard error of estimate was also calculated as: s_{YX} = ±4.9 ml kg^{-1} min^{-1}. These results are presented in Figure 3 and predicted $\dot{V}O_{2max}$ scores for each stage of the MSFT are displayed in Table 2.

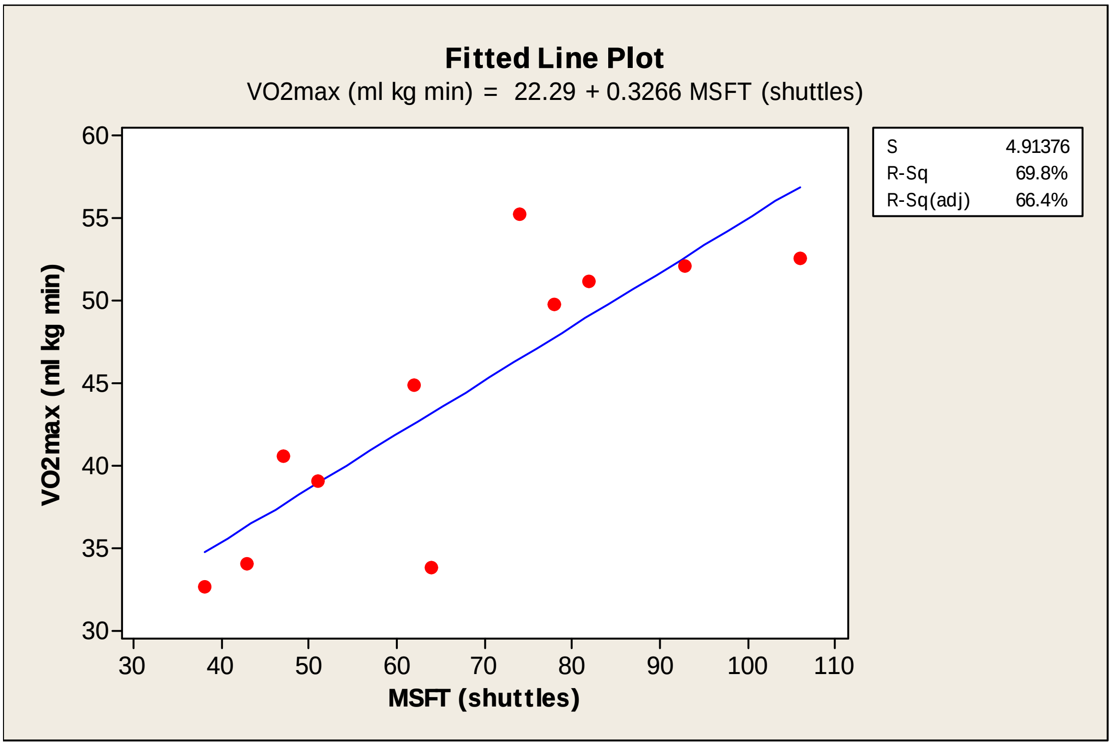

Figure 3 - Fitted line plot representing the calibration model developed in stage two of this investigation.

4.3 Stage Three

The five participants that were not included in the calibration sample were used for the purpose of cross-validation. Average scores (shuttles) of the two MSFTs completed by the subjects in stage one were calculated and used as their MSFT score for stage three ($\bar{x}$ (SD) = 70(±18) shuttles). These scores were then entered into the calibration model developed in stage two and predicted $\dot{V}O_{2max}$ scores (ml kg^{-1} min^{-1}) were calculated. Participants in the cross-validation group also performed a laboratory-based cycle $\dot{V}O_{2max}$ test ($\bar{x}$ (SD) = 44.3(±7.1) ml kg^{-1} min^{-1}. A paired *t*-test provided evidence of no significant bias between predicted and directly measured $\dot{V}O_{2max}$ scores ($\bar{x}_{diff}$ (±SD_{diff}) = -3.7 (±1.8) ml kg^{-1} min^{-1}, t = -0.99, P = 0.38). The residual errors were found to be normally distributed (A^2 = 0.53, P = 0.084) and the heteroscedasticity coefficient was close to zero (r = 0.077, P = 0.902). The bias ± the 95% LoA was therefore calculated as: -3.7 ±3.5 ml kg^{-1} min^{-1}. A Pearson's correlation between predicted and measured $\dot{V}O_{2max}$ scores showed a strong relationship (r = 0.96; P < 0.05). Figure 4 highlights the results of this stage of the investigation.

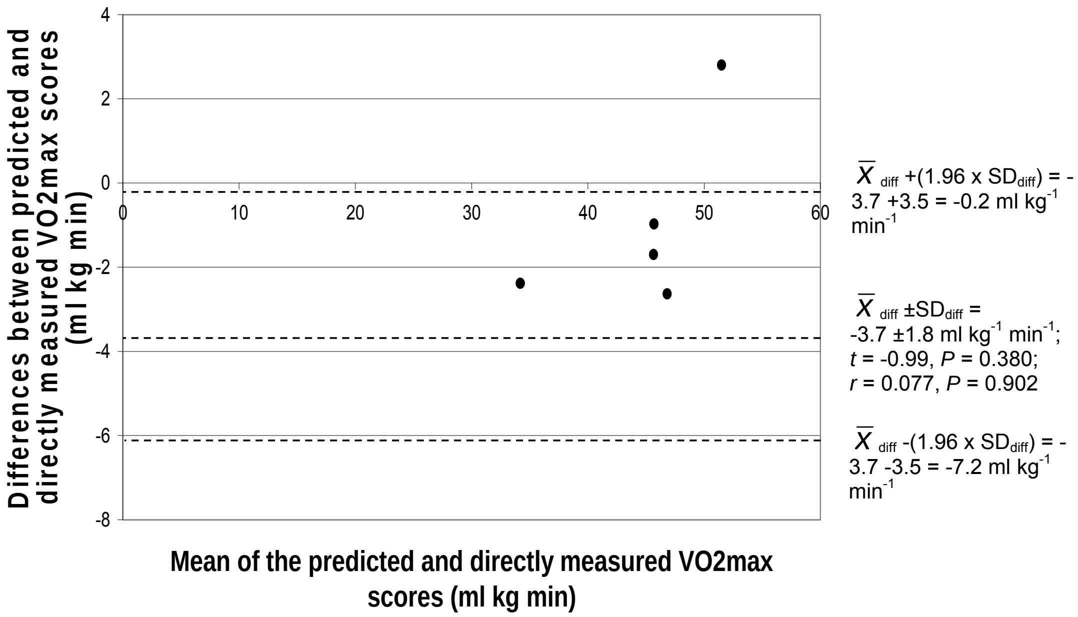

Figure 4 - Bland-Altman plot highlighting the results of the cross-validation stage of this investigation. LoA, bias and heteroscedasticity data have also been superimposed on the plot.

Table 2 - Predicted VO_{2max} (ml kg^{-1} min^{-1}) scores for each stage of the MSFT

Stage	Cumulative Shuttles	Predicted VO_{2max}
1	7	24.6
2	15	27.2
3	23	29.8
4	31	32.4
5	40	35.4
6	49	38.3
7	59	41.6
8	69	44.9
9	79	48.1
10	90	51.7
11	101	55.3
12	113	59.3
13	125	63.2
14	138	67.4
15	151	71.7
16	165	76.3
17	179	80.8
18	194	85.7
19	205	89.3
20	220	94.2

Chapter Five – Discussion and Conclusions

5.1 Results

5.1.1 Stage one

Stage one of this investigation aimed to quantify the test-retest reliability of the adapted 20m MSFT used; through utilisation of 95% LoA analysis. Other studies had used simple correlation coefficients to quantify reliability. The assessment of reliability through the use of correlation coefficients could have been misleading however. This was because the strength of correlations can be largely attributed to the range of the scores used to obtain them. A high correlation does not give an indication of agreement between tests. To put this into perspective, the correlation coefficient of MSFT test-retest scores from this investigation was compared to their LoA. The results indicated a very strong correlation (r = 0.98, P <0.05), between test and retest scores and this was comparable to other studies which also found the same strength of correlation (Léger and Lambert, 1982; Cooper *et al*., 2005). This result could have led to the interpretation of the MSFT being a repeatable and reliable test.

The 95% limits of agreement calculated (-3 ±3 shuttles), suggested that a subject’s test and retest MSFT scores could be as much as six shuttles apart (to a 95% confidence level). On average this amounted to a possible test-retest difference of up to 8.8% for the sample population used. Despite the seemingly wide LoA, it compared favourably to the LoA experienced by Lamb and Rogers (2007) of -5.3 ±16.3 shuttles.

Based on the results of the two methods of analyses; the agreement between test and retest scores was not fully represented by the high correlation between them. This was because the correlation method did not give indications of the amount of error to be expected. Therefore, the use of the LoA to quantify reliability in this investigation was justified.

For the LoA to be considered acceptable, the difference between the test-retest scores had to be narrow enough for the results of each test to be used interchangeably. Whether the LoA were narrow enough or not was 'dependant upon the measurement and the use to which it was to be put' (Bland, 2006). Lamb and Rogers (2007) considered the LoA that they experienced, to be too wide for the MSFT to be considered reliable. The LoA from the current study were under half of those experienced in the aforementioned study. Another consideration when quantifying reliability was the systematic bias found. The mean difference was -3 shuttles, which suggested a bias towards retest scores. This value was not statistically significant however. A heteroscedasticity coefficient of less than 0.2 was found and homoscedasticity was assumed as $P > 0.05$. This was important as it meant that the LoA could be expressed as 1.96 x SD_{diff}, without further transformation of data (Bland, 2006). The residual errors were normally distributed and consequently the LoA were considered to be narrow enough to consider the scores from the version of the MSFT used, to be reliable.

5.1.2 Stage two

This stage aimed to develop a calibration model for predicting $\dot{V}O_{2max}$ scores from scores on an adapted MSFT. In this case, the model was developed for a rugby specific sample group, but previous research has used similar methods of calibration modelling to predict $\dot{V}O_{2max}$ scores from MSFT scores for varying populations. The most important considerations when comparing results across studies were the relationship between $\dot{V}O_{2max}$ and MSFT scores, the common variance between test scores and the standard error of estimate. A strong correlation between test scores was essential because it determined the amount of common variance between scores and the accuracy of predicted scores.

It was suggested by Vincent (1999), that correlation coefficients must exceed r = 0.8 for the common variance between scores (coefficient of determination, R^2), to be sufficiently high (R^2 > 64%). A correlation weaker than r = 0.8 would likely have caused a common variance between test scores of less than 50%, making any findings much less significant. The current study reported a correlation coefficient of r = 0.835 and an adjusted coefficient of determination of R^2 = 66.4%. This meant that the relationship between the two test scores was sufficient enough to calculate meaningful predictions from the calibration model developed. Several studies have reported even stronger correlations of r > 0.90 (Léger and Lambert, 1982; Palickza *et al*., 1987; Ramsbottom *et al.*, 1988).

However, not all studies have found strong relationships; Léger *et al*. (1988) found a correlation between MSFT and $\dot{V}O_{2max}$ scores of only r = 0.71 and thus the test scores only had 50% common variance.

The standard error of estimate gave an indication of how much error could be expected when predicting $\dot{V}O_{2max}$ scores using the calibration model. Expressed within 95% confidence limits, the standard error of estimate gives a range either side of the predicted value that we can be 95% certain that the correct measured value lies within. Whether or not the standard error of estimate is sufficiently small enough for predictions to be of significance depends on the magnitude of the predicted scores and for what use they are intended. Results displayed a standard error of estimate for the calibration model developed in stage two of this investigation to be ±4.9 ml kg^{-1} min^{-1}. This meant that predicted and measured $\dot{V}O_{2max}$ scores could be different by as much as 9.8 ml kg^{-1} min^{-1}. This error range was considered to be very high after taking into consideration that the fact that the average $\dot{V}O_{2max}$ for subjects from the sample population was only 44.2 ml kg^{-1} min^{-1}. However, when the standard error of estimate was compared with those in studies by Léger and Lambert, (1982) and Léger *et al*. (1988), the s_{YX} value observed in the current study was surprisingly lower; by ±0.5 and ±1.0 ml kg^{-1} min^{-1}. In both cases, the MSFT was still concluded to be a valid predictor of $\dot{V}O_{2max}$ and so it was justified for the current investigation to make the same conclusion.

5.1.3 Stage three

The third stage aimed to cross-validate the calibration model developed in stage two on a random sample from the same population that developed the model. This was done by examining the agreement between predicted $\dot{V}O_{2max}$ max scores (calculated by entering MSFT scores into the calibration model), with $\dot{V}O_{2max}$ max scores determined from a laboratory-based cycle test. The 95% LoA method was utilised once again for this purpose. As previously addressed with reliability, several previous studies had chosen to assess validity through the use of criticised correlation coefficients. For the purpose of comparison between studies, correlation between predicted $\dot{V}O_{2max}$ scores and directly measured $\dot{V}O_{2max}$ scores was calculated (r = 0.96, P < 0.05). This strong linear relationship compares favourably to previous studies that had correlations lower than r = 0.8 (Léger *et al.*, 1988; St. Clair-Gibson *et al.*, 1998) and similarly to the relationship found by Léger and Gadoury (1989), which reported a correlation of r = 0.92. Validity of the scores could have been assumed with such a strong correlation coefficient, but that analysis alone does not quantify the agreement between the scores.

Scores must agree sufficiently if the tests are to be used interchangeably (Bland, 2006). The 95% LoA were calculated at ±3.5 ml kg^{-1} min^{-1} after homoscedasticity had been confirmed (r = 0.077, P = 0.902).

These limits were much narrower than those reported by Cooper *et al*. (2005) of ±6.3 ml kg^{-1} min^{-1}, who concluded that, because of the wider LoA, the agreement between predicted and directly determined $\dot{V}O_{2max}$ scores was not sufficient enough for the MSFT to be considered a valid predictor of $\dot{V}O_{2max}$. It was decided, however, that the narrower LoA calculated in the current study were sufficient for the predicted $\dot{V}O_{2max}$ scores to be considered valid. It was also interesting to note that the study by Cooper *et al.* (2005), found significant systematic bias between the predicted and measured $\dot{V}O_{2max}$ scores, which was not eliminated, even after log and antilog transformations; the bias found in the current study was not statistically significant (P = 0.38).

Another method of quantifying cross-validation was used in the study by Stickland *et al.* (2003). The method involved complex computer-intensive determination of distributions of regression weights, by applying regression solutions to opposing sample groups 10,000 times. Average regression weights were then compared to assess the cross-validity of the regression models used. For the purpose of cross-validation, this method may be more appropriate than the 95% LoA method used in the current investigation.

5.2 Sample size

Perhaps the most obvious limitation of the current study was the relatively small amount of subjects used in each stage ($n \leq 16$ for each of the three stages of investigation). This was particularly evident when compared to several previous studies which had over 70 subjects (Léger and Lambert, 1982; Ramsbottom *et al*., 1988; Léger and Gadoury, 1989). Cooper *et al.* (2004) followed a similar three-part investigation structure and therefore provided comparisons in sample numbers used for each stage of the investigation. Stage one in the current study (reliability) used 16 subjects. This compared quite similarly to Cooper *et al*. (2004) who had 20 subjects in their reliability group. Stage 2 (calibration modelling) compared less favourably in terms of subjects numbers with n = 11 compared to n = 36 in the cited study. In the current study, cross-validation (stage 3) used only five subjects. This was much lower than the amount used by Cooper *et al*. (2005) who again had 36 subjects involved in the cross-validation stage. The amount of subjects used should be as much as possible, but it has been suggested that for purposes of regression, the sample size should exceed 40 (Vincent, 1999). However, other studies have still reported significant results with samples sizes below 40 (Palickza *et al*., 1987; St. Clair-Gibson *et al.*, 1998; Lamb and Rogers, 2007).

5.3 Characteristics of the sample group

The current study used a heterogeneous sample of twelve male and four female rugby players. The use of both male and female subjects has been criticised as it widens the characteristics of the group, potentially strengthening any correlations between test scores (St. Clair-Gibson *et al*., 1998). Earlier studies that reported the valid use of the MSFT as a predictor of $\dot{V}O_{2max}$ also used both male and female subjects (Léger and Lambert, 1982; Ramsbottom *et al*., 1988; Léger and Gadoury, 1989). Specific requirements for subjects used in the current study were enforced too. Subjects had to be aged between 18 and 25 and play rugby union. Cutting the age range of the sample and selecting a specific population helped narrow the range of the covariates. Despite this, large differences in $\dot{V}O_{2max}$ had previously been reported even within rugby specific populations (Scott *et al*., 2003). This was because the aerobic requirements of specific playing positions could vary greatly.

The age range among the sample group was also minimised. Léger *et al*. (1988) reported differences in $\dot{V}O_{2max}$ between various age groups. The use of a wide age range could therefore have had an influence on the results found. The results of the current study were still valid and reliable, despite the use of a heterogeneous sample group. Correlations were also stronger than those experienced by studies that used homogeneous groups (St. Clair-Gibson *et al*., 1998).

Despite the advantages of widening the characteristics of the sample, future studies should take the following variables into account: limiting the characteristics of the sample group as much as possible by using a single gender, decreasing age range and using players from the same or similar positions. Results of such studies would give a deeper, more specific insight into the relationships between test scores; rather than the more generic results obtained from sample groups with wide ranging characteristics.

5.4 Habituation in the MSFT

For the purposes of reliability, subjects performed two MSFTs; a test and retest. On average, the sample group performed better in the retest; 68 (±19) shuttles in the retest compared to only 67 (±21) in the initial test. This resulted in a bias being observed of -3 shuttles, but this value was statistically insignificant. It was not determined whether further retests would agree more closely or not. Lamb and Rogers (2007) investigated the effect that performing more than one repeat test can have on the reliability of the MSFT. They found much closer agreement between second and third performances of MSFTs (0.8 ±18.2 shuttles), than first and second trials (-4.3 ±19.2 shuttles), or first and third trials (-5.3 ±16.3 shuttles). The systematic bias between test results was cut by over 75% between second and third trials when compared to between the first and second trials. It was therefore suggested that habituation can improve reliability greatly. However, it was still noted by Lamb and Rogers (2007) that much random error was still present in the data, despite the elimination of the systematic bias due to habituation.

Future investigations should therefore adopt multiple repeat trials in order to eliminate systematic bias evident between an initial test and single retest.

5.5 Issues in $\dot{V}O_{2max}$ testing

Several issues with the $\dot{V}O_{2max}$ test protocol used in the current study were identified. McArdle *et al*. (2000) suggest $\dot{V}O_{2max}$ test protocols should last between eight and twelve minutes. Subjects greatly exceeded this guideline in the majority of cases. This could have meant that either the intensity of the test was too low, or that the stage durations were too long. When the findings of Roffey *et al*. (2007) are considered, the three minute stage durations used in the current studies protocol, were perhaps too long and one or two minute stage durations should have been used.

Another debatable issue observed in the $\dot{V}O_{2max}$ test protocol was the use of a cycle ergometer as opposed to a treadmill. Tests should be as specific as possible to increase the common variance. As rugby union and the MSFT involve running, a treadmill protocol would be more specific to the test and the sport and therefore most appropriate. For the reasons addressed in the review of literature it was decided that a cycle protocol would be used for the current investigation. Many subjects complained of tiring leg muscles when performing the cycle $\dot{V}O_{2max}$ test. Subjects also indicated the belief that they could have continued further had the demands on their leg muscles not been as high.

McCartney *et al.* (1983) reported that the lower $\dot{V}O_{2max}$ reported when cycle $\dot{V}O_{2max}$ protocols are used, may be due to the high torque required by the leg muscles to meet the demands of the increasing load. This meant that subjects with stronger leg muscles could have potentially kept cycling for longer durations than those subjects with weaker leg muscles. This was backed up by Jasólska *et al*. (1999) who found higher maximal peak and mean power outputs in cycle ergometry, than in treadmill protocols. However, $\dot{V}O_{2max}$ scores determined by cycle ergometry were found by LeMura *et* al. (2001) to be significantly lower than those obtained via use of treadmill protocols. Hill *et al.* (2002) reported these differences in $\dot{V}O_{2max}$ to be as much as 11% lower in cycle protocols than the $\dot{V}O_{2max}$ measured using treadmill test protocols. The use of a treadmill protocol may therefore have prevented premature withdrawal / disqualification from the $\dot{V}O_{2max}$ test due to localised muscular fatigue.

Only two BASES defined criteria of $\dot{V}O_{2max}$ achievement were considered during the $\dot{V}O_{2max}$ testing (maximum heart rate and volitional exhaustion). Other criteria such as a respiratory exchange ratio above 1.15 may have provided a better guide to achieving $\dot{V}O_{2max}$. However, due to the protocol used, such calculations could not be made in real time. Equipment is available, however, that could have provided more detailed results in real time (Oxycon). Future studies should consider the use of such equipment as the simpler criteria that define $\dot{V}O_{2max}$ are not the most accurate for assessing attainment of $\dot{V}O_{2max}$.

The formulas used to calculate $\dot{V}O_{2max}$ were also simplified. More accurate results could have been obtained if the temperature of (Charles' law), pressure of (Boyle's law) and water vapour contained within the expired gases had been taken into consideration. This is because the volume of expired gases could have varied depending on those variables.

5.6 Conclusions

The results obtained from this investigation sufficiently met the original aims outlined in chapters 1 and 2. Based on the findings, it was concluded that scores from the 20m multistage fitness test, performed on a turf pitch whilst wearing studded boots, are sufficiently reliable and repeatable when the analysis performed is compared to existing literature. A calibration model for predicting $\dot{V}O_{2max}$ (ml kg^{-1} min^{-1}) from MSFT scores (shuttles completed), was created. It was based on a relatively strong, positive linear correlation and this accounted for a sufficiently high common variance between test scores. Although the standard error of estimate was perceived to be quite high, it was deemed acceptable after comparison to support literature. Cross-validation of the calibration model resulted in a good level of agreement between predicted and directly measured $\dot{V}O_{2max}$ values. Again the results were favourable when compared to similar studies. Overall, this study has found that scores obtained from an adapted multistage fitness test, performed by a sample of male and female rugby players, provide reliable and valid predictions of $\dot{V}O_{2max}$ when entered into the calibration model: $\dot{V}O_{2max}$ = 22.3 + (0.327 x MSFT shuttles completed).

List of References:

- Altman, D.G. (1999). Practical statistics for medical research. London: Chapman and Hall.
- Australian Sports Commission. (2000). Physiological tests for elite athletes. Champaign, IL: Human Kinetics.
- Aziz, A.R., Tan, F.H.Y. and Teh, K.C. (2005) A pilot study comparing two field tests with the treadmill run test in soccer players. *Journal of Sports Science and Medicine*; **4**:105-112.
- Baker, J., Ramsbottom, R. and Hazeldine, R. (1993). Maximal shuttle running over 40 m as a measure of maximal exercise performance. *British Journal of Sports Medicine*; **27**:228–32.
- Baker, J.S. and Davies, B. (2002). High intensity exercise assessment: relationships between laboratory and field measures of performance. *Journal of Science and Medicine in Sport*; **5(4)**:341-347.
- Basset, F.A. and Boulay, M.R. (2000). Specificity of treadmill and cycle ergometer tests in triathletes, runners and cyclists. *European Journal of Applied Physiology*; **81(3)**:214-221.
- Basset, F.A. and Boulay, M.R. (2003). Treadmill and cycle ergometer tests are interchangeable to monitor triathletes annual training. *Journal of Sports Science and Medicine*; **2**:110-116.
- Bland, J. and Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet.

- Bland, M. (2006). An introduction to medical statistics. 3rd edition. Oxford: Oxford University Press.
- Brewer, J., Ramsbottom, R., Williams, C. (1988). Multistage fitness test. Leeds: National Coaching Foundation.
- Bryman, A. and Cramer, D. (1996). Quantitative data analysis with Minitab. London: Routledge.
- Buchfuhrer, M.J., Hansen, J.E., Robinson, T.E., Sue, D.Y., Wasserman, K. and Whipp, B.J. (1983). Optimising the exercise protocol for cardiopulmonary assessment. *Journal of Applied Physiology*; **55**: 558-564.
- Carter, H., Jones, A.M., Barstow, T.J., Burnley, M., Williams, C.A. and Doust, J.H. (2000). Oxygen uptake kinetics in treadmill running and cycle ergometry: a comparison. *Journal of Applied Physiology*; **89(3)**:899-907.
- Cooper, S.M., Baker, J.S., Eaton, Z.E. and Matthews, N. (2004). A simple multistage field test for the prediction of anaerobic capacity in female games players. *British Journal of Sports Medicine*; **38**:784-789.
- Cooper, S.M., Baker, J.S., Tong, R.J., Roberts, E. and Hanford, M. (2005). The repeatability and the criterion related validity of the 20m multistage fitness test as a predictor of maximal oxygen uptake in active young men. *British Journal of Sports Medicine*; **39**:19-27.

- Dabonneville, M., Berthon, P., Vaslin, P. and Fellmann, N. (2003). The 5 min running field test and retest reliability on trained men and women. *European Journal of Applied Physiology*; **88**:353-360.
- Gabbett, T., Kelly, J. and Pezet, T. (2007). Relationship between physical fitness and playing ability in rugby league players. *Journal of Strength and Conditioning Research*; **21(4)**:1126-1133.
- Grant, S., Corbett, K., Amjad, A.M., Wilson, J. and Aitchison, T. (1995). A comparison of methods of predicting maximum oxygen uptake. *British Journal of Sports Medicine*; **29**:147-152.
- Hazeldine, R. and McNab, T. (1991). Fit for rugby. London: Kingswood Press.
- Hill, D.W., Davey, K.M. and Stevens, E.C. (2002). Maximal accumulated O_2 deficit in running and cycling. *Canadian Journal of Applied Physiology*; **27(5)**:463-478.
- Hill, D.W., Halcomb, J.N. and Stevens, E.C. (2003). Oxygen uptake kinetics during severe intensity running and cycling. *European Journal of Applied Physiology*; **89(6)**:612-618.
- Jaskólska, A., Goossens, P., Veenstra, B., Jaskólski, A. and Skinner, J.S. (1999). Comparison of treadmill and cycle ergometer measurements of force-velocity relationships and power output. *International Journal of Sports Medicine*; **20(3)**:192-197.
- Karpman, V.L. (1987). Cardiovascular System and Physical Exercise. Boca Raton, FL: CRC Press, Inc.

- Kline, G., Porcari, J., Freedson, P.S., Ward, A., McCarron, R.F., Ross, J., Rippe, J.M. and Hintermeister, R. (1987). Estimation of $\dot{V}O_{2max}$ from a one-mile track walk, gender, age and body weight. *Medicine and Science in Sport Exercise*; **19**:253–259.
- Lamb, K.L. and Rogers, L. (2007). A re-appraisal of the reliability of the 20m multi-stage shuttle run test. *European Journal of Applied Physiology*; **100**:287-292.
- Léger, L. and Gadoury, C. (1989). Validity of the 20m shuttle run test to predict $\dot{V}O_{2max}$. *Canadian Journal of Applied Sport Sciences*; **14**:21-26.
- Léger, L., and Lambert, J. (1982). A maximal multistage 20-m shuttle run test to predict $\dot{V}O_{2max}$. *European Journal of Applied Physiology*; **49**:1–12.
- Léger, L., Mercier, D., Gadoury, C. and Lambert, J. (1988). The multistage 20 metre shuttle run test for aerobic fitness. *Journal of Sports Sciences*; **6**:93–101.
- Léger, L., Seliger, V. and Blassard, L. (1980). Backward extrapolation of $\dot{V}O_{2max}$ values from the O2 recovery curve. *Medicine and Science in Sport and Exercise*; **12(1)**:24-27.
- LeMura, L.M., von Duvillard, S.P., Cohen, S.L., Root, C.J., Chelland, S.A., Andreacci, J., Hoover, J. and Weatherford, J. (2001). Treadmill and cycle ergometry testing in 5- to 6-year-old children. *European Journal of Applied Physiology*; **85(5)**:472-478.
- Lohman, T.G. (1988). Anthropometric standardisation reference manual. Champaign, IL: Human Kinetics.

- Luger, D. and Pook, P. (2004). Complete conditioning for rugby. Champaign, IL: Human Kinetics.
- Maud, P.J. and Foster, C. (2006). Physiological assessment of human fitness. 2nd edition. Champaign, IL: Human Kinetics.
- McArdle, W.D., Katch, F.I. and Katch, V.L. (2000). Essentials of exercise physiology. 2nd edition. Baltimore, MD: Williams and Wilkins.
- McCartney, N., Heigenhauser, G.J.F. and Jones, N.L. (1983). Power output and fatigue of human muscle in maximal cycling exercise. *Journal of Applied Physiology*; **55(1)**: 218-225.
- Nijsten, S.M., van Elmpt, W.J., Jacobs, M., Mijnheer, B.J., Dekker, A.L., Lambin, P. and Minken, A.W. (2007). A global calibration model for a-Si EPIDs used for transit dosimetry. *Medical Physics*; **34(10)**:3872-84.
- Palickza, V., Nichols, A. and Boreham, C. (1987). A multistage shuttle run as a predictor of running performance and maximal oxygen uptake in adults. *British Journal of Sports Medicine*; **21(4)**:163–164.
- Rampinini, E., Bishop, D., Marcora., S.M., Ferrari Bravo, D., Sassi, R. and Impellizzeri, F.M. (2006). Validity of simple field tests as indicators of match related physical performance in top level professional soccer players. *International Journal of Sports Medicine*; **28(3)**:228-235.
- Ramsbottom, R., Brewer, J. and Williams, C. (1988). A progressive shuttle run test to estimate maximal oxygen uptake. *British Journal of Sports Medicine*; **22(4)**:141–144.

- Roffey, D.M., Byrne, N.M. and Hills, A.P. (2007). Effect of stage duration on physiological variables commonly used to determine maximum aerobic performance during cycle ergometry. *Journal of Sports Sciences*; **25(12)**:1325-1335.
- Scott, A.C., Roe, N., Coats, A.J. and Piepoli, M.F. (2003). Aerobic exercise physiology in a professional rugby union team. *International Journal of Cardiology*; **87(2-3)**:173-7.
- Shephard, R.J. (1994). Aerobic fitness and health. Champaign, IL: Human Kinetics.
- Siegler, J., Robergs, R. and Weingart, H. (2006). The application of soccer performance testing protocols to the non-elite player. *Journal of Sports Medicine and Physical Fitness*; **46(1)**:44-51.
- St. Clair-Gibson, A., Broomhead, S., Lambert, M.I. and Hawley, J.A. (1998). Prediction of maximal oxygen uptake from a 20m shuttle run as measured directly in runners and squash players. *Journal of Sports Sciences*; **16**:331-335.
- Stickland, M.K., Petersen, S.R. and Bouffard, M. (2003). Prediction of maximal aerobic power from the 20m multi-stage shuttle run test. *Canadian Journal of Applied Physiology*; **28(2)**:272-282.
- Tanner, J.M. (1949). Fallacy of per-weight and per-surface area standards and their relation to spurious correlation. *Journal of Applied Physiology*; **2**: 1-15.

- Tharp, G., Newhouse, R., Thorland, W.G., Johnson, G.O. and Uffleman, L. (1984). Comparison of sprint and run times with performance on the Wingate Anaerobic Test. *Research Quarterly*; **56**:73–76.
- The British Association of Sport and Exercise Sciences. (2007). Sport and exercise physiology testing guidelines. Abingdon: Routledge.
- Thomas, J.R., Nelson, J.K. and Silverman, S.J. (2005). Research methods in physical activity. 5th edition. Champaign, IL: Human Kinetics.
- Tomczak, C.R., Warburton, D.E., Riess, K.J., Jendzjowsky, N.G., Liang, Y., Bhambhani, Y., Haennel, R.G. and Haykowsky M.J. (2007). A prediction model for estimating pulmonary oxygen uptake during the 6-minute walk test in organ transplant recipients. *Transplant Proceedings*; **39(10)**:3313-3316.
- Uth, N., Sørensen, H., Overgaard, K. and Pedersen, P.K. (2004). Estimation of $\dot{V}O_{2max}$ from the ratio between HR_{max} and HR_{rest} – the heart rate ratio method. *European Journal of Applied Physiology*; **91**:111-115.
- Valderrama, P., Braga, J.W. and Poppi, R.J. (2007). Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. *Journal of Agriculture and Food Chemistry*; **55(21)**:8331-8338.

- Van Schuylenbergh, R., Eynde, B.V. and Hespel, P. (2004). Prediction of sprint triathlon performance from laboratory tests. *European Journal of Applied Physiology*; **91(1)**:94-99.
- Vehrs, P.R., George, J.D., Fellingham, G.W., Plowman, S.A. and Dustman-Allen, K. (2007). Submaximal treadmill exercise test to predict $\dot{V}O_{2max}$ in fit adults. *Measurement in Physical Education and Exercise Science*; **11(2)**:61-72.
- Vincent, W.J. (1999). Statistics in kinesiology. 2nd edition. Champaign, IL: Human Kinetics.
- Wilmore, J.H. and Costill, D.L. (2004). Physiology of sport and exercise. 3rd edition. Champaign, IL: Human Kinetics.

Appendix A

UWIC
School of Sport

Participant Consent Form

UWIC Ethics Protocol Number:
Participant Name:
Study: The 20m multistage fitness test as a predictor of VO_{2max} in rugby players.
Researcher: Carl Corcoran

Participants to complete the following section by putting their initials in each box.

1. I have been informed of the nature and procedures involved in the testing and understand the requirements of the study. I have had the opportunity to consider the information and ask questions about the tests. Any questions raised have been answered to my satisfaction. ☐
2. I understand that my participation is voluntary and that I am free to withdraw from testing at any point, without need for reason and without compromising any professional relationship or my legal rights. ☐
3. I understand that sections of any research notes and data collected during the study may be reviewed by responsible individuals from UWIC for monitoring purposes, where it is relevant to my taking part in this study. I give permission for these people to have access to my records. ☐
4. Having considered all previous information, I agree to take part in this study. ☐

__________________________ __________________
Signature of Participant | Date

____
Name of Person Taking Consent

__________________________ __________________
Signature of person taking consent | Date

UWIC
School of Sport

Physical Activity Readiness Questionnaire (PAR-Q)

Please circle appropriate responses to the following questions:

1. Has your doctor ever said that you have a heart condition and that you should only perform physical activity recommended by a doctor?
Yes / No

2. Do you feel pains in the chest when you do physical activity?
Yes / No

3. In the past month have you had any chest pain whilst not performing physical activity?
Yes / No

4. Do you lose your balance because of dizziness or do you ever lose consciousness?
Yes / No

5. Do you have a bone or joint problem that could be made worse by physical activity?
Yes / No

6. Is your doctor currently prescribing you with drugs for high blood pressure or a heart condition?
Yes / No

7. Do you know of any other reason why you should not participate in the required physical activity?
Yes / No

If you have answered yes to any of the above questions, please add details below. Similarly, if there are any situations that will prevent you from exercising write them here.

____________________________ ________________

Signed Date

UWIC **RISK ASSESSMENT (RA99)** (V3/07) *Page 1 - (Hazards)*

School / Unit and Area:		**Assessment Number:**	
Risk Assessment undertaken by: Recommended to be 2 or more people			
Description of the work activity being assessed:			
Persons Affected:	Staff ☐ Students ☐ Others ☐		
Details of Others:			

HAZARD IDENTIFICATION Please provide details of the hazards associated with the area or task. **EXAMPLES INCLUDE:** Working at height, Manual Handling, Electricity, Fire, Noise, Contact with moving parts of machinery, Dust etc		**RISK RATING - <u>without</u> Controls** The Risk Rating **(RR)** and Degree of Risk are determined by multiplying the Severity **(S)** of injury by the Likelihood **(L)** of occurrence. Please see UWIC Risk Rating Matrix for details			
		S	**L**	**RR**	**Degree of Risk**
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
Example - 1. Electric Shock (office)		4	3	12	Unacceptable

Once all potential hazards have been identified and a Risk Rating has been applied, please go to page 2 and provide details of the control measures required to reduce the risk to an acceptable level.

UWIC RISK ASSESSMENT (RA99) *Page 2 – (Controls)*

CONTROLS TO BE APPLIED Examples Include: Elimination, Substitution for something less hazardous, Barriers or fixed guards, standard operating procedures and personnel protective equipment		Date Applied	RISK RATING - with Controls			
			S	L	RR	Degree of Risk
1						
2						
3						
4						
5						
6						
1.	**Examples of possible controls:** All appliances are to be PAT tested. Any new items are to be reported to estates. Users to undertake visual checks prior to use. Damaged equipment to be removed from use.	07/06/07	4	1	4	Moderate

Date of First Assessment:		Review Date of overall Assessment:	

Appendix B

This table gives the age and physical characteristics of each subject, along with test scores and heart rate data

Subject	Age	Par-Q	Informed Consent	Stature	Body Mass	Resting HR	HRmax	Test max HR	% of HRmax achieved
1	21.3	x	x	188.5	112.8	56	199	200	100.7
2	18.4	x	x	192.1	81.9	73	202	202	100.2
3	18.7	x	x	184.7	69.8	80	201	189	93.9
4	18.6	x	x	182.6	73.6	74	201	187	92.9
5	21.5	x	x	189.3	64.5	83	199	201	101.3
6	22.3	x	x	176.5	79.5	67	198	185	93.6
7	21.8	x	x	176.5	65.2	79	198	183	92.3
8	21.0	x	x	181.4	74.2	67	199	196	98.5
9	20.4	x	x	188.0	91.7	56	200	184	92.2
10	21.3	x	x	159.1	58.1	69	199	202	101.7
11	21.1	x	x	181.9	82.7	50	199	194	97.5
12	21.2	x	x	183.8	79.0	62	199	187	94.1
13	21.2	x	x	187.1	84.6	62	199	189	95.1
14	21.7	x	x	179.6	74.4	74	198	200	100.9
15	20.8	x	x	174.2	74.4	72	199	188	94.4
16	21.0	x	x	170.3	58.7	85	199	182	91.5
Mean	**20.8**			**181.0**	**76.6**	**69.3**	**199.2**	**191.8**	**96.3**
St Dev	**1.2**			**8.3**	**13.4**	**10.2**	**1.2**	**7.3**	**3.7**
Calibration - mean	**20.6**			**181.9**	**77.6**				
St Dev	**1.4**			**9.1**	**15.1**				
Cross Validation - mean	**21.2**			**179.0**	**74.2**				
St Dev	**0.3**			**6.9**	**9.6**				

MSFT 1	Shuttles	MSFT 2	Shuttles	Average shuttles	VO2max	Date of VO2max test
6.1	50	6.3	52	51	39.1	12/02/2008
7.3	62	7.3	62	62	44.9	06/02/2008
8.1	70	8.8	77	74	55.3	06/02/2008
5.3	43	6.1	50	47	40.6	06/02/2008
8.6	76	9.1	80	78	49.8	12/02/2008
5.1	41	5.4	44	43	34.1	11/02/2008
4.3	35	5.0	40	38	32.7	11/02/2008
9.3	82	9.3	82	82	51.2	12/02/2008
10.1	95	10.5	91	93	52.1	12/02/2008
7.4	63	7.5	64	64	33.8	05/02/2008
11.8	109	11.2	103	106	52.6	05/02/2008
9.3	82	8.6	75	79	45.5	05/02/2008
9.5	84	9.6	85	85	52.9	05/02/2008
8.3	73	8.4	74	74	44.8	06/02/2008
8.3	73	8.3	73	73	45.2	06/02/2008
5.1	41	4.6	38	40	33.0	11/02/2008
	67.4		**68.1**	**67.8**	**44.2**	
	21.2		**19.1**	**20.1**	**7.8**	
	Calibration Group		**Mean**	**66.9**	**44.2**	
			St Dev	**21.8**	**8.5**	
	CV Group		**Mean**	**69.8**	**44.3**	
			StDev	**17.6**	**7.1**	

Minitab worksheet used in stage one of this investigation

MSFT 1	MSFT 2	Means	Residuals	Absolute Differences
50	52	51	-2	2
62	62	62	0	0
70	77	73.5	-7	7
43	50	46.5	-7	7
76	80	78	-4	4
41	44	42.5	-3	3
35	40	37.5	-5	5
82	82	82	0	0
95	91	93	4	4
63	64	63.5	-1	1
109	103	106	6	6
82	75	78.5	7	7
84	85	84.5	-1	1
73	74	73.5	-1	1
73	73	73	0	0
41	38	39.5	3	3

Minitab worksheet used for stages two and three of the investigation

Calibration group VO_{2max} (ml kg min)	Calibration group MSFT (shuttles)	Cross-Validation VO_{2max} (ml kg min)	Cross-Validation MSFT (shuttles)
39.1	51	45.5	79
44.9	62	52.9	85
55.3	74	44.8	74
40.6	47	45.2	73
49.8	78	33	40
34.1	43		
32.7	38		
51.2	82		
52.1	93		
33.8	64		
52.6	106		

Predicted VO_{2max}	Residuals	Means	Absolute Differences
48.1	-2.6	46.8	2.6
50.1	2.8	51.5	2.8
46.5	-1.7	45.6	1.7
46.2	-1.0	45.7	1.0
35.4	-2.4	34.2	2.4

This table shows the calculated VO_{2max} scores for each subject

along with their body weight and expired gas during each stage of the lab test

		Stage	FEO_2	$FECO_2$	*V*E (L)	Adjusted	VO_{2max} L min^{-1}	VO_{2max} ml kg^{-1} min^{-1}
Subject	13	1	18.2	2.9	36	37	1.01	11.9
Mass	84.6	2	17.2	3.5	42.9	43.9	1.64	19.4
		3	17	4	57.8	58.8	2.31	27.3
		4	17.8	3.2	82.2	83.2	2.60	30.8
		5	17.8	3.5	110	111	3.47	41.1
		6	17.9	3.5	146.8	147.8	4.48	52.9
	12	1	16.2	4.3	17.9	18.9	0.89	11.3
	79	2	16	4.9	28.9	29.9	1.47	18.7
		3	16	5	34.7	35.7	1.76	22.3
		4	16.6	4.5	57.8	58.8	2.55	32.2
		5	17	4.4	91.1	92.1	3.62	45.8
		6	18.1	3.4	126.1	127.1	3.60	45.5
	10	1	18.9	1.8	28	29	0.59	10.1
	58.1	2	18.3	2.5	31.8	32.8	0.86	14.8
		3	18.2	2.6	34.8	35.8	0.98	16.8
		4	18.2	2.7	37.7	38.7	1.06	18.2
		5	17.5	3.6	41.1	42.1	1.44	24.9
		6	17.4	4	54.6	55.6	1.96	33.8

11	1	16.6	4.4	35.4	36.4	1.58	19.1
82.7	2	15.8	4.9	38.8	39.8	2.04	24.7
	3	15.8	5	47.5	48.5	2.49	30.1
	4	16	4.8	56.9	57.9	2.85	34.5
	5	16.4	4.6	71.4	72.4	3.28	39.7
	6	16.5	4.1	85.8	86.8	3.85	46.5
	7	17	4.3	109.6	110.6	4.35	52.6
2	1	17.9	3.7	60.1	61.1	1.85	22.6
81.9	2	18.3	3.4	78.2	79.2	2.08	25.4
	3	17.6	3.8	68.6	69.6	2.32	28.3
	4	17.9	3.7	104.6	105.6	3.20	39.1
	5	18.3	3.3	138.9	139.9	3.68	44.9
3	1	17.5	3.8	39.1	40.1	1.38	20.0
68.8	2	17.1	4.2	42.1	43.1	1.65	24.0
	3	16.7	4.5	49.7	50.7	2.14	31.2
	4	17	4.3	69.2	70.2	2.76	40.1
	5	17.3	4.2	91.3	92.3	3.35	48.7
	6	17.7	3.8	116.8	117.8	3.80	55.3
4	1	17.6	3.4	36.9	37.9	1.26	17.1
73.6	2	17.3	3.8	45.1	46.1	1.67	22.7
	3	17.3	4	60.1	61.1	2.22	30.1
	4	17.7	3.6	78.9	79.9	2.58	35.1
	5	17.9	3.6	97.6	98.6	2.99	40.6
15	1	16.3	5.2	31.9	32.9	1.52	20.5
74.4	2	17.4	4.2	51.2	52.2	1.84	24.8
	3	17.6	3.9	59.2	60.2	2.00	26.9
	4	17.5	4	83.4	84.4	2.89	38.9
	5	17.8	3.8	106.4	107.4	3.36	45.2

14	1	17.4	3.6	42.3	43.3	1.53	20.5
74.4	2	16.9	4.1	47.4	48.4	1.95	26.2
	3	16.5	4.4	52	53	2.35	31.6
	4	17.1	4.1	68.3	69.3	2.65	35.7
	5	17.1	4.2	85.6	86.6	3.32	44.6
	6	17.3	4.3	90.9	91.9	3.34	44.8
16	1	17.8	3.5	29.5	30.5	0.95	16.3
58.7	2	18	3.4	39.4	40.4	1.18	20.2
	3	17.9	3.6	52.7	53.7	1.63	27.7
	4	17.9	3.5	62.9	63.9	1.94	33.0
6	1	16.4	4.1	21.3	22.3	1.01	12.7
79.5	2	16.2	4.4	27.8	28.8	1.36	17.1
	3	16.4	4.5	36.5	37.5	1.70	21.4
	4	16.8	4.3	48.4	49.4	2.04	25.7
	5	17	4.2	59.1	60.1	2.36	29.7
	6	17.6	3.8	80.5	81.5	2.71	34.1
7	1	17	4	22.7	23.7	0.93	14.3
65.2	2	17.3	4	34.6	35.6	1.29	19.8
	3	17.4	4	44.7	45.7	1.61	24.7
	4	17.5	3.9	61.2	62.2	2.13	32.7
9	1	17.2	3.9	38.7	39.7	1.48	16.1
91.7	2	16.9	4.1	46	47	1.89	20.7
	3	16.7	4.3	53.3	54.3	2.30	25.0
	4	16.7	4.3	62.5	63.5	2.69	29.3
	5	16.4	4.5	69.5	70.5	3.19	34.8
	6	17	4.2	91.9	92.9	3.65	39.8
	7	17.1	4	106.1	107.1	4.10	44.7

	8	17.4	3.9	134.3	135.3	4.78	52.1
8	1	16.3	4.4	28.8	29.8	1.38	18.6
74.2	2	16	4.8	35.3	36.3	1.79	24.1
	3	16.3	4.7	49.5	50.5	2.34	31.5
	4	16.8	4.4	63.7	64.7	2.67	36.0
	5	17.1	4.3	80.9	81.9	3.14	42.3
	6	17.8	3.7	120.3	121.3	3.80	51.2
5	1	17	4.1	41.1	42.1	1.65	25.7
64.5	2	17.5	4	59.3	60.3	2.07	32.1
	3	17.3	4.1	70.4	71.4	2.59	40.2
	4	17.8	3.7	101.7	102.7	3.21	49.8
1	1	17.7	3.8	61.9	62.9	2.03	18.0
112.8	2	17.4	4	60.3	61.3	2.16	19.2
	3	17.5	4	73.3	74.3	2.55	22.6
	4	17.4	3.9	89.3	90.3	3.19	28.3
	5	17.6	3.8	124.5	125.5	4.18	37.0
	6	18.3	3.2	166.9	167.9	4.42	39.1