1989

Wingate test : reliability and learning effect

Paul A. Schumann
Ithaca College

WINGATE TEST: RELIABILITY AND LEARNING EFFECT

By

Paul A. Schumann

An Abstract

of a thesis submitted in partial fulfillment

of the requirements for the degree of

Master of Science in the Division

of Health, Physical Education,

and Recreation at

Ithaca College

September 1989

Thesis Advisor: Dr. G. A. Sforzo

ABSTRACT

The purposes of this study were to determine the trial-to-trial reliability of the Wingate Anaerobic Test (WAT), to investigate whether any learning effect occurs, and if so, to establish a trend-free trial schedule to derive reliable data. Male subjects (18-24 years of age) were randomly assigned to one of two testing schedules, either daily (n = 12) or twice-a-day (n = 13). The daily subjects performed the WAT once a day for 6 consecutive days, whereas the twice-a-day subjects performed the WAT twice a day for 3 consecutive days, with an hour recovery between trials on each day. The traditional WAT protocol was used in each exercise session to determine peak power (PP), average power (AP), and power decline (PD). Two-way (Groups X Trials) ANOVAs indicated there were no significant differences for the trial-to-trial scores between the daily and twice-a-day groups for any of the anaerobic parameters. There were significant differences in both groups for the trial-to-trial scores in PP and AP, but not in PD. The observed differences indicated that a learning effect was occurring between Trial 1 and Trial 2. The results support the ideas that multiple testing will increase the reliability of the WAT and that the learning effect

can be reduced by following a trend-free trial schedule. The schedule of choice depends on the anaerobic parameter to be tested and the degree of reliability the researcher will accept. Reliability estimates and recommendations for trial schedules have been made for each parameter. In addition, other more practical trial schedules have been recommended for investigation.

WINGATE TEST: RELIABILITY AND LEARNING EFFECT

A Thesis Presented to the Faculty of
the Division of Health, Physical
Education, and Recreation at
Ithaca College

In Partial Fulfillment of the
Requirements for the Degree
Master of Science

---

by

Paul A. Schumann

September 1989

Ithaca College
Division of Health, Physical Education, and Recreation
Ithaca, New York

CERTIFICATE OF APPROVAL

MASTER OF SCIENCE THESIS

This is to certify that the Master of Science Thesis of

Paul A. Schumann

submitted in partial fulfillment of the requirements for the degree of Master of Science in the Division of Health, Physical Education, and Recreation at Ithaca College has been approved.

Thesis Advisor:

Committee Member:

Candidate:

Chairman, Graduate Programs in Physical Education:

Dean of Graduate Studies:

Date: 8/19/89

ACKNOWLEDGMENTS

The investigator wishes to express his sincere appreciation to everyone involved in the completion of this study:

1. Dr. Gary A. Sforzo, my thesis advisor, for all his advice, support, and patience throughout the investigation.

2. Dr. Patricia A. Frye, my committee member, for all her help and guidance with the statistical analysis, as well as her valued suggestions and corrections throughout the investigation.

3. Dr. A. Craig Fisher, my graduate advisor, for his valued suggestions throughout my studies.

4. John W. Murray, for all his helpful advice all through my years at Ithaca College.

5. Joel Lamson, Bob Panariello, all my friends, and also the subjects who devoted time and energy in this investigation.

6. Scott Harrell, for his support and understanding throughout the investigation.

7. Bob, Beth, and Patti, my brother and sisters, and best friends, for all their devoted support throughout the investigation and in all my endeavors.

8. Most of all, my parents, Robert and Rita Schumann, for believing in me and supporting me all

through the years at Ithaca College and in all that I have attempted to accomplish throughout my life.

DEDICATION

This study has been dedicated in loving memory of Anthony J. Hanson, whose love and commitment to his family will shine bright forever.

TABLE OF CONTENTS

APPENDIXES

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Considerable research effort is directed toward the measurement and interpretation of maximal aerobic power. Different test protocols for the measurement of maximum oxygen consumption (VO_2 max) are utilized, and the reliability and validity of these tests are well known. However, far less attention is focused on the determination and meaning of tests of anaerobic functioning. This is quite disconcerting, because more sporting and individual athletic events rely upon anaerobic than aerobic performance. At this time, there appears to be no agreement as to which measurement tool designates the most valid estimate of anaerobic functioning.

In retrospect, with its origin as a measure of anaerobic power and capacity in 1974, the Wingate Anaerobic Test (WAT) has gained wide acceptance as an important exercise protocol for the determination of anaerobic performance. The advantages of this test are that it is simple to administer and it takes little time to perform. There have been a number of studies concerning modifications of the optimal load and/or duration for a variety of populations, as well as studies of WAT results under various testing conditions. In related research, there have been several studies examining the optimal test characteristics for maximal anaerobic work, in addition to evaluations of tests of anaerobic power and all-out anaerobic capacity tests on cycle ergometers. It could be said that considerable information gathered regarding optimal Wingate testing conditions has made it the leading test of anaerobic performance.

Although the past research warrants special consideration for refining the measurement and interpretation of anaerobic power parameters via the WAT, there has been limited, if any, research concerning the correct number of trials to elicit the most reliable and optimal WAT test results. Furthermore, it is unknown if a learning effect occurs with repeated WAT administration, and if so, whether the learning effect can be controlled with a trend-free trial schedule. Therefore, because the WAT is a potentially important measure for research and functional evaluation, it was imperative to examine these factors.

Statement of Problem

The purposes of this study were to (a) establish the reliability of the WAT, (b) determine if a learning effect occurred with repeated WAT administration, and

(c) if a learning effect was present, establish a trend-free trial schedule for the WAT.

Scope of Problem

An attempt was made to determine the trial-to-trial reliability of the WAT, to investigate whether any learning effect may occur, and if so, to establish a trend-free trial schedule to derive reliable data. For these purposes two multiple WAT protocols were established. The subjects for this study were 25 healthy male subjects, ages 18 to 24 years. Subjects were randomly divided into either a daily or a twice-a-day testing schedule. The daily subjects performed the WAT once a day for 6 consecutive days, whereas the twice-a-day subjects performed the WAT twice a day for 3 consecutive days, with an hour recovery between testing on each day. The testing was performed on a Monark cycle ergometer fitted with racing-type toe clips. Pedal revolutions were counted electro-mechanically and recorded every 5 s to calculate three anaerobic parameters, namely peak power (PP), average power (AP), and power decline (PD).

Hypotheses of Study

The following null hypotheses concerning the trial-to-trial reliability of the WAT, occurrence of a learning effect, and the establishment of a trend-free testing schedule were identified:

Ho_1: There will be no significant differences when comparing the trial-to-trial scores for the daily and twice-a-day schedules within the groups or across the groups, for any of the anaerobic parameters.

Ho_2: There will be no differences in the reliabilities produced by different trial combinations for any of the three anaerobic parameters determined from the WAT.

Definition of Terms

The following terms that were used in this study are herein defined:

1. Anaerobic capacity: The ability to persist at the maintenance or repetition of strenuous muscular contractions that rely substantially upon anaerobic mechanisms of energy supply (Lamb, 1984).

2. Anaerobic power: The maximal rate at which energy can be produced or work can be performed without a significant contribution of aerobic energy production (Lamb, 1984).

3. Average power (AP): The maximal capacity to produce energy anaerobically by breaking down ATP, creatine phosphate, and glycogen. This is estimated with the WAT.

4. Peak power (PP): The power created by the splitting of ATP and creatine phosphate. This is estimated with the WAT.

5. Power decline (PD): An index of fatigue rate calculated as a percentage of peak power. This is estimated with the WAT.

6. Wingate Anaerobic Test (WAT): A 30-s all-out cycling test to determine maximal anaerobic power and maximal anaerobic capacity.

Assumptions of Study

The following assumptions concerning the study were made:

1. Subjects clearly understood all the requirements that were necessary to correctly perform the WAT, and performed to the best of their ability during each testing session.

2. Sufficient recovery was allowed between subsequent testing trials.

3. Subjects did not change their exercise habits throughout the course of the study.

4. The WAT is a representative measure of anaerobic power and anaerobic capacity, and the techniques employed in this study yielded accurate WAT scores.

5. A six-trial testing schedule is sufficient to determine a trend-free trial schedule.

Delimitations of Study

The following were the delimitations of the study:

1. Only 25 male subjects, 18 to 24 years of age, were tested.

2. This study was limited to two testing schedules, a daily group and a twice-a-day group.

3. The WAT protocol was used to derive the anaerobic parameter data.

4. The testing was performed on a standard Monark cycle ergometer, without instant load capability.

5. Subject population was limited to individuals weighing 170 lb (77.3 kg) or below.

Limitations of Study

The following limitations exist for this study:

1. The results of this study only apply to males, ages 18 to 24 years, weighing less than 170 lb (77.3 kg), equivalent to those used in this study.

2. The results of this study apply only to two testing schedules, a daily or a twice-a-day schedule.

3. The results of this study apply only to anaerobic capacity parameters derived from the WAT.

4. The results of this study can be applied only to testing performed on a standard Monark cycle

ergometer, and results may be different with an instant load cycle.

Chapter 2

REVIEW OF LITERATURE

The measurement of work and mechanical power on a cycle ergometer is a useful method for evaluating maximal power and capacity derived from anaerobic metabolism. The development of procedures for measuring the ability of muscle to generate power during high-intensity exercise has received considerable attention in recent years. Procedures have included simple field tests, treadmill running, stair-climbing, vertical jumping, isokinetic knee extension, and cycle pedaling (Patton & Duggan, 1987). Although no individual test has gained the popularity equal to the determination of VO_2 max as a measure of aerobic power, the WAT has received considerable attention as the preferred measure of anaerobic performance. There have been a number of studies concerning modifications of the optimal load, crank length, and duration for a variety of populations, as well as WAT results under various testing conditions. In related research, there have been several studies examining the optimal test characteristics for maximal anaerobic work, in addition to evaluations of tests of anaerobic power and all-out anaerobic tests on cycle ergometers.

This chapter will review the literature pertinent to this study. Specifically, this chapter will include the following sections: (a) factors affecting WAT performance, (b) reliability and validity, (c) learning effect, and (d) summary.

Factors Affecting WAT Performance

The WAT was developed in the early 1970s at the Wingate Institute for Sport and Physical Education, Israel. The WAT is a 30-s supramaximal test that can be performed by either cycling or arm cranking (Inbar & Bar-Or, 1986). The three parameters calculated from the WAT are PP, AP, and PD, and they are said to be dependent upon anaerobic ability. Factors that affect WAT administration and performance would include test procedure modifications and physiological factors.

Test Procedure Modifications

Although crucial in proper test administration, load selection had been originally based on only a limited pilot study. The resistance used for leg exercise was 0.075 kg/kg of body weight (BW) for Monark cycle ergometers. While this resistance has been reported to elicit the greatest power outputs, these data were obtained from a specific population, namely, young subjects (Patton, Murphy, & Frederick, 1985). Therefore, Dotan and Bar-Or (1983) investigated the optimal loads (OL) for eliciting maximal power outputs (PO) in leg and arm WAT testing in 18 adult female and 17 adult male subjects. Five randomized resistance loads ranging from 2.43 to 5.39 J/pedal rev/kg BW for the legs, and from 1.96 to 3.92 for the arms, were utilized. The resulting OLs were 5.04 and 5.13 in the leg and 2.82 and 3.52 in the arm tests for the women and men, respectively.

Also based on the results, it was evident that PP and AP could not be optimized through a single test. Dotan and Bar-Or (1983) suggested that while the WAT appears to be rather insensitive to moderate variation in load assignment, improved results could be obtained by using the stated OL guidelines that may be modified according to an individual's body build, body composition, and anaerobic fitness level. It was also recommended that stated OL values be used as group or population load-selection guidelines that could be altered somewhat to accommodate interindividual differences (Dotan & Bar-Or).

Patton et al. (1985) used 19 male subjects in their study of maximal power outputs with workloads ranging between 3.23 and 6.76 J/pedal rev/kg BW. Data obtained at the 5.59 resistance load and that of the traditional WAT setting of 4.41 were used for statistical comparison. Both PP and AP were significantly higher at the 5.59 workload than at the 4.41 workload. PD was also significantly higher at 5.59 as a result of the greater force applied and the lower mean pedal revolutions obtained (Patton et al.). At resistances above 5.88, some subjects had difficulty in maintaining contact with the saddle throughout the test. Based on the results, Patton et al. suggested that considerable variability occurred among the male subjects in resistances eliciting maximal PO values for leg exercise. Also, it was found that those loads can not be reliably predicted using BW or thigh volume. The WAT workload based solely on total BW does not take into account the active muscle mass. Evans and Quinney (1981) have developed an optimal resistance using body mass and leg volume. However, this determination of optimal resistance has received considerable opposition from Katch (1974) and Patton et al., in which they found that leg volume was of little predictive importance. Therefore, Patton et al. concluded that the resistance load continue to be assigned according to the subject's BW, but consideration be given to increasing the workload from 4.41 that is presently used by many investigators.

It is apparent that there is considerable controversy regarding the standardization of the optimal testing workload for the WAT, as well as the correct method to predict the optimal workload. Therefore, based on the current research, Vandewalle, Peres, and Monod (1987) have made the following observations: for adult men, it appears to be better to use a braking force equal to around 0.095 kg/kg BW. The value of 0.086 kg/kg BW, determined by the Wingate Institute, seems to be optimal for women, whereas 0.075 kg/kg BW should be sufficient for children.

According to Inbar, Dotan, Trousil, and Dvir (1983), no research prior to 1983 questioned the utilization of the traditional 17.5 cm crank length (CL) for the various types of cycling performance. Inbar et al. tested 13 male subjects to determine if the optimal CL for maximizing anaerobic performance was different than that previously used in WAT administration. The Monark ergometer was modified with a crank-slider assembly which permitted continuous CL adjustment. Five evenly distributed CLs were administered, ranging from 12.5 cm to 22.5 cm. For both indices the highest mean values occurred at the 15 cm CL. The parabolic maximum occurred between the actual maximum at 15 cm and the traditional CL of 17.5

cm (16.4 and 16.6 cm for AP and PP, respectively). Inbar et al. suggested that the optimal CL determined for both AP and PP was only marginally shorter than the traditional 17.5 cm CL. The subjects were a generally homogeneous population, and it was concluded that the optimal CL was not considerably different from the traditionally used CL. Additionally, when the subjects were divided into short-, intermediate-, and long-legged groups, Inbar et al. (1983) suggested that the optimal CL for longer or shorter individuals should deviate from the 17.5 cm CL by 1 cm for approximately every 6.3 cm difference in leg length. Based on the findings, Inbar et al. concluded that CL optimization should be both task and subject dependent.

Frederick et al. (1983) reported modifications of the Monark cycle ergometer that allowed for the instantaneous application of resistance, as well as the methods used to calibrate the Monark ergometer for instantaneous application of resistance. Further investigation concerning whether the traditional resistance application or the instantaneous application of resistance elicits superior results with WAT administration is warranted. Additionally, LaVoie, Brayne, and Barrett (1984) reported that the use of toe stirrups in anaerobic power testing, performed on a cycle ergometer, appears crucial for maximal WAT performance. When the ergometer was equipped with toe stirrups, the subjects could use both the flexors and extensors of both legs throughout the entire pedal revolution. LaVoie et al. concluded that the utilization of toe stirrups elicited higher power outputs.

Physiological Factors

Environmental heat stress and its effect on human performance have been extensively studied. Hot or humid climates have been found to be detrimental to prolonged or aerobic type work (Saltin, 1964). However, information concerning the effects of climatic heat stress on all-out anaerobic performance has received limited attention. To determine climatic effects on the performance of the WAT, Dotan and Bar-Or (1980) tested 28 children in three different climates. The three testing groups were as follows: a neutral group, in which the climate was 22 to 33 °C with 55% to 60% relative humidity (RH); a hot-dry climate of 38 to 39 °C and 25% to 30% RH; and a warm-humid group with a climate of 30 °C and 85% to 90% RH. Average power, relative to body weight, in the boys after the warm-humid exposure was higher than after the hot-dry exposure. No other significant differences were found for either the boys or the girls in any of the climates.

The general conclusion made was that performance of the WAT was not affected by various climatic conditions. The practical implication of the study is that it is not necessary to strictly monitor climatic conditions whenever the WAT is to be performed outside the laboratory. These findings were in agreement with Bar-Or, Dotan, and Inbar (1977), who also found that the performance of this test was independent of short exposures to different climates. Dotan and Bar-Or (1980) suggested that more information is still necessary regarding anaerobic performance in more extreme heat stress and when heat exposure time is longer than 45 min. Also, how anaerobic performance is affected by cold climates is not well known.

Dehydration by means of exercise, heat diuretics, semistarvation, or a combination of these is common among competitors in weight class sports. The effects of these practices have demonstrated a reduced aerobic work capacity (Allen, Smith, & Miller, 1977; Herbert & Ribisl, 1972; Palmer, 1968). The effects of these practices on anaerobic performance is not well documented. Therefore, Jacobs (1980) used 11 collegiate wrestlers to investigate the effects of progressive, acute, thermal dehydration to a 5% weight loss on performance of the WAT. The subjects performed the WAT prior to and after each of the following mean weight losses: 2%, 4%, and 5%. Average power and peak power were not significantly changed by acute, thermal dehydration to a 5% weight loss. However, both absolute PP and PP per kg BW increased due to the weight loss. Jacobs proposes that anaerobic performance may not be negatively affected to the degree that aerobic performance is by passive, thermal dehydration to a 5% body weght loss. However, negative physiological effects may result from dehydration practices though performance levels are maintained.

Performance during the WAT of PP, total work performed (TW), and PD, have been reported to be positively related to the fast-twitch composition of the vastus lateralis muscle for male physical education students and trained runners (Bar-Or et al., 1980: Inbar, Kaiser, & Tesch, 1981). These findings are consistent with the data recorded by Citerrio and Agostoni (1984), showing a selective activation of fast-twitch fibers of the quadricep muscles at high cycling speeds. This was further investigated by Froese and Houston (1987), when they examined the relationship between muscle fiber composition of the vastus lateralis muscle and performance during the WAT. The subjects consisted of 18 females and 12 males. The indices studied included PP, TW, PD, and posttest blood lactate concentration. The absolute values for PP, TW, and PD were significantly greater for the males than for the females, whereas lactate concentration was not significantly higher, although found to be 22% greater for the males. Because BW and leg volume were used to determine the resistance settings, PP and TW were calculated per unit of body mass and per unit of the combined volume of both legs for the subjects. These relative values for the males exceeded those of the females by 17% to 22%. For the male subjects, PP, TW, PD, and posttest blood lactate concentration were all significantly correlated with both the percentage of fast-twitch (%FT) fibers and the percentage relative area of fast-twitch (%FTA) fibers. However, no significant correlations were present for any of the measures within either %FT and %FTA for the female subjects. Based on the results, Froese and Houston emphasized the importance of muscle morphology for the determination of short-term anaerobic performance for untrained males.

The nonsignificant relationship between indices of muscle morphology and anaerobic performance for the female subjects warrants further consideration. Vandewalle, Peres, Heller, and Monod (1985) proposed that women tend to have a larger proportion of body fat and a lower muscle mass than men, and therefore, the workload based on Evans and Quinney's (1981) leg volume protocol probably resulted in force settings that were too high in terms of active muscle mass. Also, Bell & Jacobs (1986) indicated that women subjects may have a slower rate of isometric force development than male subjects. Froese and Houston (1987) suggested that women may have more difficulty producing a high power output, especially a burst of power as needed when beginning the WAT. Additionally, Jacobs, Tesch, Bar-Or, Karlsson, and Dotan (1983) have reported lower lactate concentrations in the vastus lateralis muscle of female than in male subjects after 10 and 30 s of performance on the WAT. The above mentioned results that tend to suggest that female subjects have a lower anaerobic power and capacity, coupled with the findings of Froese and Houston, substantiate the conclusion of Dotan and Bar-Or (1983) that selection of a workload for the WAT must also take into consideration the sex of the subject as well as his/her anaerobic fitness level.

There have been conflicting data concerning the aerobic component of the WAT. Inbar, Dotan, and Bar-Or (1976) determined the aerobic component to be 13% with a net oxygen consumption of 1.07 L over the duration of the WAT. Jacobs (1980) proposed a similar aerobic contribution of 13.2%, whereas, Stevens and Wilson (1986) reported a 49.3% aerobic component with a net oxygen consumption of 1.67 L. The large discrepancies need clarification before assuming that the WAT is a valid test to evaluate anaerobic performance (Kavanaugh & Jacobs, 1988). Kavanaugh and Jacobs tested five male subjects to determine the aerobic contribution to the WAT. The mean net VO_2 corresponded more directly to the results of Inbar et al. and Jacobs than with the findings of Stevens and Wilson. Kavanaugh and Jacobs suggested from the results that subjects with a higher VO_2 max may be able to perform a greater proportion of the WAT energy requirement with aerobic metabolisn than subjects with a lower VO_2 max. Hence, the presence of a significant aerobic component to the WAT would enable the subjects with higher VO_2 max values to elicit the larger relative aerobic contributions (Kavanaugh & Jacobs).

Although the WAT has been a useful tool to measure anaerobic performance, the traditional WAT does not enable the subject to work using arms and legs simultaneously. Therefore, Koutedakis and Sharp (1986) examined a modified WAT for measuring anaerobic work of the upper body. The Double Arm Anaerobic Work Test (DAAWT) was utilized to test 24 male junior rowers. The DAAWT is a computerized double-arm modification of the WAT, in which work is performed by using both arms to turn a bar connectiong the cranks of two Monark cycle ergometers (Sharp, Koutedakis, Slater, Harris, & Katsikas, 1986). The optimal resistance was determined to be 8% of the subject's BW. Based on the data from the study, Koutedakis and Sharp proposed that the modified version of the traditional WAT was a reasonably good method for obtaining information regarding the upper body performance of junior rowers.

Reliability and Validity

The reliability of anaerobic tests is generally determined by means of the test-retest coefficient of correlation. The validity of the test-retest coefficient of correlation as a measure of reliability is debatable (Vandewalle et al., 1987). Wilmore (1968) stated that while the test-retest correlations are of considerable interest, they should be interpreted with caution as the test-retest coefficients of correlation are highly dependent on the size of the population and the degree of interindividual variability. Therefore, a high coefficient would be expected if the interindividual variability is high, even when the sample size is small, and the converse would also be expected to be true. Thus, the validity of anaerobic tests is often difficult to measure because a test must first be reliable before it can be valid (Vandewalle et al.).

Test-retest correlation cofficients for the WAT have been reported for various anaerobic indices. Patton et al. (1985) calculated the reliability for PP, AP, and PD for two resistance loads. Correlation coefficients for PP and AP ranged from .91 to .93 for resistances of 4.41 and 5.59 J/rev/kg BW, respectively. The reliability coefficients for PD were considerably lower at .74 and .43 for both resistances.

Dotan and Bar-Or (1980) investigated the reliability of the WAT in nonstandardized environments. Correlation coefficients of power output, using male and female subjects, ranged from .94 to .73 for the three climatic conditions. Comparing the test-retest correlation coefficients of .95 to .97 proposed by Bar-Or et al. (1977), Dotan and Bar-Or's values appear lower. There were likely two reasons for the lower values. Approximately 2 weeks separated consecutive tests, which was long enough to elicit changes in the subjects' overall state of condition. Also, the subject distribution was much more limited than that of a normal population. Dotan and Bar-Or concluded that the practicality and reliability of the WAT is well maintained in field situations where climate cannot be carefully controlled.

The WAT has been found to be reliable with a test-retest reliability ranging from .95 to .97 in various groups of children, adolescents, and young adults (Bar-Or et al., 1977). Several investigations were made to validate the 30-s anaerobic test against accepted criteria of anaerobic capacity: its correlation with 300-m running time was -.85, with maximal O_2 debt was .86, and with 25-m swimming time was -.87 to -.90. Bar-Or et al. proposed that the 30-s anaerobic test was a simple and reliable test which fatigued the anaerobic pathways and could act as an indicator of anaerobic performance in various events.

Learning Effect

The goal of most researchers conducting an experimental study is to determine a highly reliable testing measure which maximizes the mean performance of the subjects used in the study. Because subjects vary as to their eagerness to perform at peak capacity, several trials of the criterion are recommended (Baumgartner & Jackson, 1970). This requires the researcher to determine the proper number of trials to administer in order to obtain highly reliable results, as well as maximum performance. Unfortunately, the correct number of trials to administer is often unknown, or the past recommended number of trials does not consider the previous experience of the subjects.

Because of these uncertainties, Baumgartner and Jackson (1970) tested five groups of subjects on a variety of physical performance tasks, in which multiple trials were administered for each task. Optimal measurement schedules of consecutive test trials, designed to elicit a reliable criterion measure and maximize the mean performance of the groups, were obtained for each task. It was established, when administering tests involving jumping, that administering five to six trials and calculating the criterion score from the later trials was a superior measurement schedule than administering three trials, because the initial trials acted as warm up and/or practice.

For agility tests, the last two to three trials of the six-trial measurement schedule were better than the first two to three trials, again attributed to a warm-up and/or learning effect. Typically with agility tests, multiple-trial schedules are required so that the subjects can learn the motor task. Baumgartner and Jackson (1970) proposed that in order to maximize the mean performance of a group of subjects and to obtain acceptable reliability for a multiple-trial physical-performance task, more test trials must be utilized than are presently being administered. Additionally, in research, where precision of measurement is so vital, determining a reliable criterion of maximum performance is necessary, even if it results in administering many test trials.

Katch, McArdle, Pechar, and Perrine (1974) administered three 8-s trials with a 1-min rest period between trials, in which the movement pattern of the legs during pedaling was identical to that of a standard Monark ergometer. The testing was conducted to determine the degree of variability in force-output scores on a given day, as well as on consecutive days. Reliability coefficients were calculated between the average force scores on Days 1 and 2, and between the corresponding trials on the 2 days. There were significant increases in the force output on Day 1 between Trials 2 and 3. Also on Day 2, the force-output scores between Trials 2 and 3 were significantly higher ($p < .01$). Additionally, when the Day 1 trials were compared to the corresponding Day 2 trials, the Day 2 trials were significantly higher ($p < .01$). The average force scores for Day 2 were also significantly higher than the average force scores of Day 1. Katch et al. suggested that the results identified a small trial-to-trial and day-to-day practice effect. Because the Day 2 trial scores were significantly higher than the Day 1 trial scores, it was proposed that preliminary practice trials be administered before any trial scores are used in subsequent data analysis. The same practice or learning effect in exerting maximal force on consecutive days or trials has previously been reported for static contractions (Laycoe & Marteniuk, 1971). Katch et al. concluded that, if the absolute values of force output are to be compared between different populations of subjects, it would be essential to establish that a plateau had been reached in the measured force-output scores.

Laycoe and Marteniuk (1971) found that learning may play a role in apparent strength increases as a result of skill improvement that occurs with practice of the complex task. Inbar and Bar-Or (1975), testing 7- to 9-year-old boys with an anaerobic test similar to the WAT, found that learning may have still been a factor even after a 1-week separation between two consecutive testing sessions. Wilmore (1968) also identified a significant learning effect for work output and riding time between Day 1 and Day 2. In a related study, Spodick (1975) submitted that investigators of physiologic and pathologic exercise responses have often determined that the first exercise session was a learning experience, so that typically every subject performed significantly better the second session. Therefore, an exercise study should be based on at least the third exercise session rather than the second.

The remainder of this section will focus on research implying that a learning effect may be occurring with multiple-trial testing. Dotan and Bar-Or (1983) suggested that there was a need for muscular coordination linked to repetitive, high-frequency motion. Also, maximizing the short-duration power required for WAT performance was impeded by the circular motion of the pedals, which affected the nature of force application and increased the amount of skill and coordination necessary for the given motion-sequence frequency. Jacobs (1980) found that the WAT demanded a high degree of neuromuscular coordination for the high speed of cycling during the initial seconds of the WAT performance. Vandewalle et al. (1987) proposed that high performances and improvements in the WAT, although related to several other factors, could be explained by limiting factors that are not specific to anaerobic metabolism. Inbar and Bar-Or (1975) found that warm up positively affected the subject's performance of the WAT. Vandewalle et al. (1985), for an anaerobic test derived from the WAT, used the first two testing sessions for warm-up and learning purposes. Finally, Sharp et al. (1986), conducting a modified WAT, allowed a few minutes for each subject to practice the criterion task before the actual testing was performed.

Summary

Considering recent WAT research, it appears that load selection was somewhat insensitive to moderate variations, although improved results could be obtained through modifications according to the subject's body build, body composition, and anaerobic fitness level. Also, it was proposed that the resistance load continue to be assigned according to the subject's BW, but consideration be given to increasing the workload from the 4.41 J/pedal revolution/kg BW that is presently used by many investigators. Additionally, optimal CL determined for both AP and PP has been found to be only marginally shorter than the traditional 17.5 cm CL. Therefore, based on these findings, CL optimization should be both task and subject dependent. Also, it has not yet been determined whether the traditional resistance application or the instantaneous application elicits superior results with WAT administration. Finally, it has been reported that the utilization of toe stirrups when performing anaerobic testing on a cycle ergometer elicited higher power outputs, and therefore appears to be crucial for maximal WAT performance.

It has been demonstrated that it is not necessary to strictly monitor climatic conditions whenever the WAT is to be performed outside the laboratory. Also, anaerobic performance may not be negatively affected to the degree that aerobic performance is by passive, thermal dehydration to a 5% BW loss. It has been emphasized that muscle morphology is of importance for the determination of short-term anaerobic performance. Although conflicting data exist, the presence of a significant aerobic component to the WAT would enable subjects with higher VO_2 max values to elicit the greater WAT scores due to a larger aerobic contribution. It has been revealed that a modified

WAT, for measuring anaerobic work of the upper body, was a reasonably good method for obtaining information regarding upper body anaerobic performance.

The WAT has been found to be a reliable and valid indicator of short-distance running and swimming. Warm up has been demonstrated to favorably affect WAT performance. Additionally, based on WAT research and related studies, a learning effect appears to play a role in increased performance with multiple-trial testing. The specific degree in which a learning effect may increase multiple-trial WAT performance warrants further investigation.

Chapter 3

METHODS AND PROCEDURES

The methods in this study are explained in this chapter. It includes the following sections: (a) selection of subjects, (b) methods of data collection, (c) treatment of data, and (d) summary.

Selection of Subjects

The subjects, who volunteered for the study, were 30 males, 18 to 24 years of age, 5 of whom withdrew before the end of the study. The 25 participants who completed the study were either residents of Ithaca, NY, or students attending either Cornell University or Ithaca College. The subjects were recruited by the investigator through personal contact. Prior to any data collection, all subjects signed an informed consent form that explained the purposes and methods of the study (see Appendix A). Also, a medical health questionnaire was completed before any testing was conducted, to insure that no testing contraindications were present (see Appendix B).

Methods of Data Collection

WAT Procedure Modifications

Participants were randomly assigned to either a daily or a twice-a-day six-trial testing schedule. The six-trial testing schedule was based on a study conducted by Baumgartner and Jackson (1970), as well as a pilot research project performed by the investigator. Baumgartner and Jackson suggested that, when administering tests involving jumping to high school boys or college men, administering five to six trials and computing the criterion score from the later trials was a better measurement schedule than administering three trials, because the initial trials could be acting as practice. The investigator's pilot research project consisted of 5 subjects who performed the WAT once a day for 4 consecutive days. It appeared from the results that learning may be occurring during the initial trial of the multiple testing WAT. Although not conclusive, it was deemed necessary to extend the study to a six-trial testing schedule, in an attempt to identify the occurrence of a learning effect and establish a trend-free trial schedule.

The daily group performed the WAT once a day for 6 consecutive days, whereas the twice-a-day group was administered the WAT twice a day for 3 consecutive days, with an hour recovery between trials on each day. The hour recovery between trials on a given day has been supported by several authors (Dotan & Bar-Or, 1983; Jacobs et al., 1983; Lamb, 1984; Mero, 1988). Lamb stated that 45 to 60 min should be allowed for recovery before the WAT is repeated. Approximately 1 hour for proper recovery was also suggested by Jacobs et al. Dotan and Bar-Or utilized a 30-min recovery period when defining the optimal loads for eliciting maximal power outputs in the legs and arms for the 30-s WAT. A 30-min recovery period was also used during a modified WAT conducted by Mero. Evans and Quinney (1981), using a modified WAT, administered two trials each day, separated by 20 min of zero-load pedaling, walking, and rest. Additionally, LaVoie et al. (1984) utilized a daily testing session consisting of two 30-s maximal cycling performances separated by a 30-min rest period.

The hour recovery period was also substantiated by a pilot study, in which blood lactates were measured through recovery as an indicator of subsequent test readiness. In that pilot study, five male subjects, aged 18 to 21 years, were administered the WAT, and blood lactates were collected both before and after exercise for a 1- hour period. In all five cases, the subjects' blood lactates returned to a resting level within the hour.

WAT Protocol

Five subjects withdrew from the present study, which left 12 subjects for the daily group and 13 subjects in the twice-a-day group. Whether the participants performed the WAT using the daily or twice-a-day schedule, all were tested utilizing the traditional WAT protocol as follows: On the day of testing, the subjects were weighed to determine the workload at which they would be pedaling. Their weights were then multiplied by 0.083 kiloponds/kg to give the testing workload in kiloponds (kp), to be applied on the Monark cycle ergometer.

The subjects then cycled for 4 min at an intensity sufficient to raise their heart rates to 150 to 160 beats per minute. The cycling was interspersed with three all-out bursts for 4 to 8 s each. To insure that all subjects received the identical procedure explanation, each subject was given a written description of the testing methods (see Appendix C). The written description was administered following the initial warm up, but prior to the initiation of each subsequent testing session. Following the warm-up session, there was a 3-min rest interval to allow for proper recovery. Following the rest period, the subjects were seated on the ergometer with their feet fastened to the pedals by means of racing-type toe clips. At this time, seat height was determined and recorded for reproduction in subsequent tests.

At the command, "Start," the subjects began cranking as fast as they could against the ergometer's inertial resistance only (no load). The full, predetermined workload was applied within 3 to 4 s, once the inertial resistance was overcome. Pedal revolution count started at that instant and lasted precisely 30 s. Subjects were instructed to avoid pacing and to maintain an all-out effort throughout the test. Pedal revolutions were counted electro-mechanically and were manually recorded by the researcher every 5 s. Following the exercise session, there was a cool down to minimize the risk of a subject fainting. The subjects were encouraged to continue pedaling at a light workload for 2 to 3 min after the completion of the test (Lamb, 1984).

Treatment of Data

The analysis of data in this study encompassed intraclass reliability estimates, a two-way (Groups X Trials) analysis of variance (ANOVA) for simple main effects, and a Tukey test. Descriptive statistics for the PP, AP, and PD parameters for each group were calculated.

Work (in W) was determined by multiplying the subject's determined workload (kp) by the number of revolutions in 30 s, and multiplying this value by the conversion factor of 11.765 (Lamb, 1984). The values for PP, AP, and PD were calculated as follows:

1. PP was determined as the greatest power in a 5-s period, and is supposed to represent maximal anaerobic power. This corresponds to the power created by the splitting of adenonosine triphosphate (ATP) and creatine phosphate (CP).

2. AP was the mean power for the six periods of 5 s each, and represents a measure of anaerobic capacity. Presumably, this corresponds to the maximal capacity to produce energy anaerobically by breaking down ATP, CP, and glycogen.

3. To determine PD, the lowest power in 5 s was subtracted from the PP score, divided by PP, and then multiplied by 100 to represent the PD percentage, which is an index of the fatigue rate (Lamb, 1984; Vandewalle et al., 1987).

Intraclass reliability estimates were calculated (Safrit, 1976) for each possible combination of two, three, four, five, or six consecutive trials. To study whether the reliability of the WAT could be increased with repeated trials, the intraclass reliability coefficients were compared visually. These trial combinations were also utilized to determine whether the daily or twice-a-day schedule elicited more reliable results for each of the three anaerobic parameters.

The trial-to-trial scores for both groups were analyzed using a two-way ANOVA for simple main effects on each anaerobic parameter, to determine if significant differences were present across or within the two testing schedules. $\underline{F}$ ratios beyond the .05 level of probability were accepted as significant. In the event that significant differences did exist, a Tukey test was used to identify where these differences occurred, thus supporting or refuting the premise of this study, which stated that learning occurs during the initial trial of a multiple-trial WAT. These analyses also helped answer the question of whether the learning effect could be controlled with a trend-free trial schedule.

Summary

In this chapter, the methods of this study were detailed, including the selection of subjects, methods of data collection, and treatment of data. It was through these methods that the reliability of the multiple-trial WAT was studied. In addition, it was investigated whether a learning effect exists with repeated WAT administration, and if so, could the learning effect be controlled by a trend-free testing schedule. Finally, with these methods it was possible to determine if the daily or twice-a-day schedule was superior in eliciting optimal reliability results.

Chapter 4

ANALYSIS OF DATA

This study was conducted to determine the trial-to-trial reliability of the WAT, to investigate whether any learning effect may occur, and if so, to establish a trend-free trial schedule to derive reliable data from WAT testing. Also, there was a statistical comparison between the two testing schedules to determine if one was superior in eliciting optimal WAT results. This was accomplished by analyzing intraclass reliability coefficient estimates, as well as trial-to-trial scores in daily and twice-a-day WAT testing schedules for each of three anaerobic parameters, PP, AP, and PD. The .05 level of significance was used for all statistical analyses. Sections in this chapter include the following: (a) descriptive statistics, (b) simple main effects and a Tukey test, (c) reliability coefficient estimates, and (d) summary.

Descriptive Statistics

The 25 subjects who completed the study varied in age from 18-24 years. The mean age of the group was 20.0 $\pm$ 1.9 years.

The mean values for the daily and twice-a-day trial scores are contained in Table 1. It is rather

Table 1

Mean Values for Daily and Twice-a-day Trial Scores

Parameter	Trials					
	1	2	3	4	5	6
PP (W)						
Daily	582.47*[a]	671.68	690.01	689.91	656.39	607.28*[a]
Twice-a-day	576.85*[b]	624.81	629.06	638.22	628.62	639.21
AP (W)						
Daily	474.51*[b]	519.56	533.68	538.01	530.93	514.78
Twice-a-day	491.40*[c]	505.49	516.73	517.51	518.97	525.90
PD (%)						
Daily	35.85	39.42	38.03	37.17	35.00	31.71
Twice-a-day	30.62	33.08	33.47	31.93	30.58	31.59

Note. Daily n = 12 and Twice-a-day n = 13.

[a]Tukey test indicated Trials 1 and 6 had significantly lower scores than Trials 2-5. [b]Tukey test indicated Trial 1 had significantly lower scores than Trials 2-6. [c]Tukey test indicated Trial 1 had significantly lower scores than Trials 5 and 6.

*$p < .05$.

apparent that, for each anaerobic parameter, the trial scores increased following the initial trial. Also, the daily scores, as a whole, appear to be greater than the twice-a-day scores when comparing individual trials across the groups.

Simple Main Effects and Tukey Test

Trial-to-trial scores for both groups were analyzed with two-way (Groups X Trials) ANOVAs for simple main effects, as well as a Tukey test to identify where any significant differences occurred. PP, AP, and PD data were analyzed individually for the two groups. These data are best inspected by examining the figures that are described in the subsequent paragraphs.

Peak Power

Figure 1 represents a graphic illustration of the means for PP and indicates the statistical analysis. There was a significant interaction of the group and trial variables, evident in the crossing of the graphs between Trials 5 and 6. The interaction indicated that the daily group pattern of mean scores was not the same as the twice-a-day group pattern. The significant interaction warranted further consideration as to whether this was caused by significant differences across the trial-to-trial scores for both groups or by

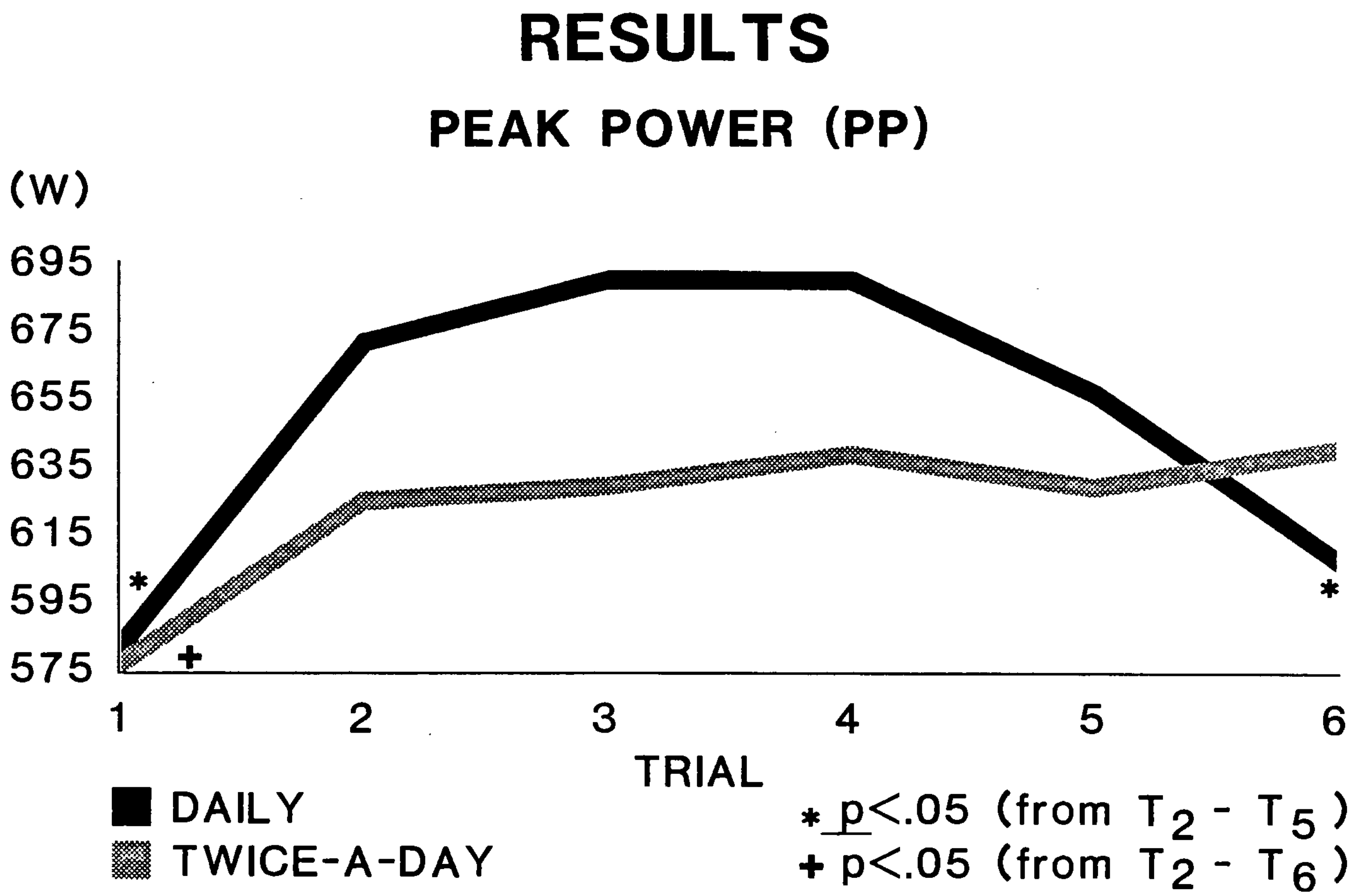

Figure 1. PP mean values for daily and twice-a-day scores for a multiple-trial WAT.

differences occurring within trial-to-trial scores for each group separately.

Comparing individual trials from the daily scores to the twice-a-day scores, there were no significant differences at any of the six trials. For example, Trial 2 scores for the daily group were not significantly different from the Trial 2 scores for the twice-a-day group. Although not considerably different, the daily means for Trials 2 through 5 were 7% higher than the means for the same trials for the twice-a-day group.

Analyzing simple main effects for each group separately, there were significant differences in the trial-to-trial scores for both the daily and twice-a-day schedules. Using a Tukey test to identify where these differences occurred, it was found that Trials 1 and 6 were significantly lower than Trials 2 through 5 for the daily group (see asterisks on Figure 1). As for the twice-a-day group, Trial 1 was significantly lower than Trials 2 through 6 (see cross on Figure 1). The significantly low initial trial may indicate the occurrence of a learning effect between Trials 1 and 2.

Average Power

Analyzing the AP parameter, graphically illustrated in Figure 2, there was a significant interaction of group and trial variables, evident in the crossing of the graphs between Trials 1 and 2 and also between Trials 5 and 6. The interaction was again evidence that the daily group's pattern of means was different from the twice-a-day group's pattern. Whether this interaction occurred because of significant differences across the trial-to-trial scores for both groups or because of differences within the trial-to-trial scores for each group separately was determined using an ANOVA for simple main effects and a Tukey test.

There were no significant differences at any of the six trials when examining the individual trials for the daily versus twice-a-day scores. Therefore, means for the daily group were not considerably different from the identical means for the twice-a-day group, despite the fact that the daily scores for Trials 2 through 5 were 3% higher than the same trials for the twice-a-day group. There were, however, significant differences between the trial-to-trial scores for both groups when analyzed separately, using an ANOVA for simple main effects. A Tukey test revealed Trial 1 was

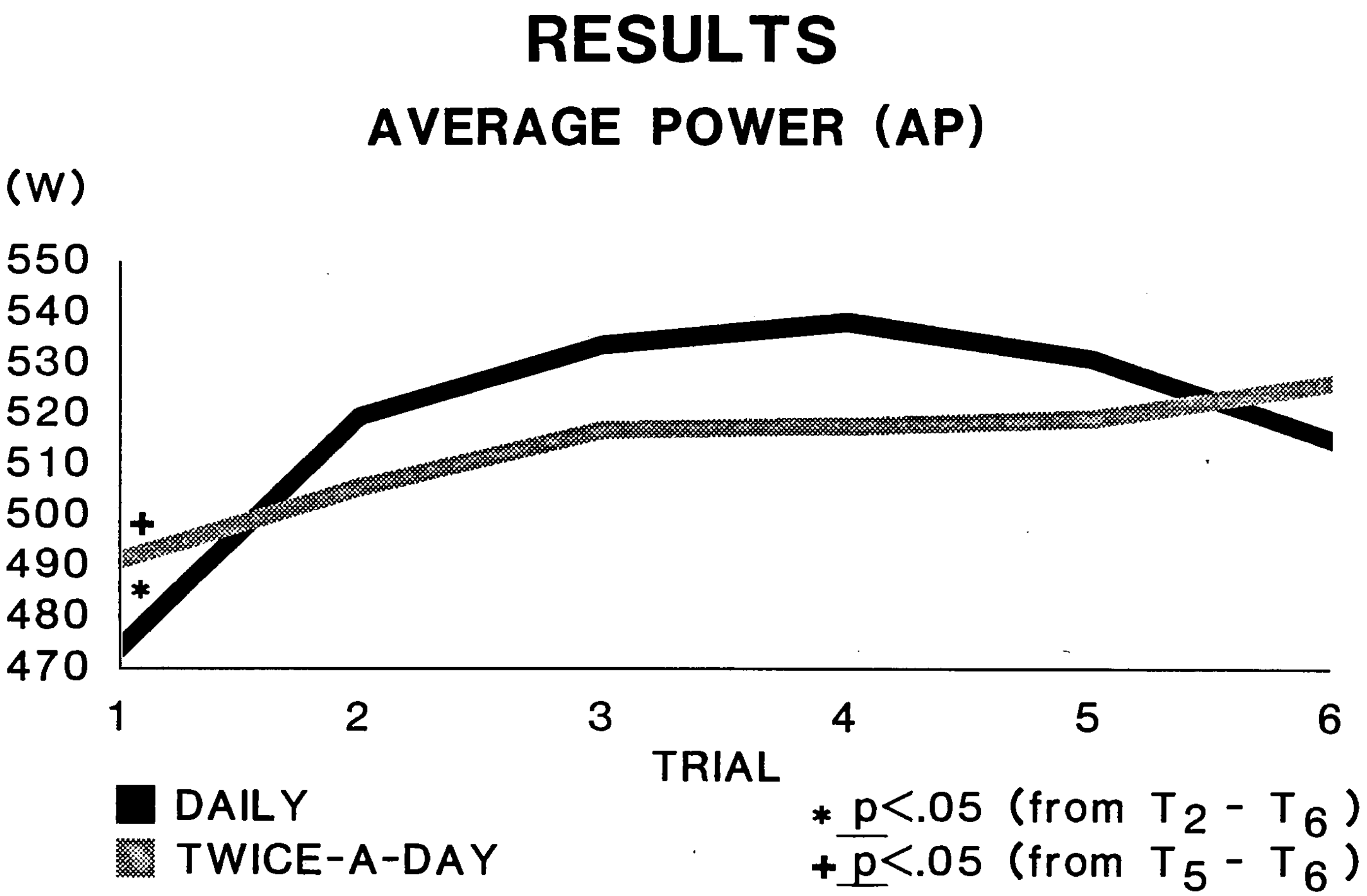

Figure 2. AP mean values for daily and twice-a-day scores for a multiple-trial WAT.

significantly lower than Trials 2 through 6 for the daily group (see asterisk on Figure 2), and Trial 1 was significantly lower than Trials 5 and 6 for the twice-a-day group (see cross on Figure 2). As with PP, the significantly low initial trial may be evidence for the occurrence of a learning effect between Trials 1 and 2.

Power Decline

Finally, there was no statistical evidence for any significant interaction of the group and trial variables for the PD parameter. Moreover, there were no significant differences when comparing the trial-to-trial scores across the groups or within the groups.

Reliability Estimates

Maximum intraclass reliability estimates for various trial combinations are depicted in Table 2. Represented in boldface are the maximum reliabilities obtained for each group for PP, AP, and PD. Appearing below the boldface maximum reliabilities are the corresponding two-trial reliabilities, in which only the initial two trials were used for statistical analysis (see Table 2).

The daily and twice-a-day reliability estimates were not considerably different. For example, the

Table 2

Maximum Reliability Estimates

	Daily Schedule		Twice-a-day Schedule	
Parameter	Trials	R	Trials	R
PP	**3-4[a]**	**.963**	**3-6[b]**	**.969**
	1-2	.762	1-2	.884
AP	**3-4[a]**	**.975**	**2-6[b]**	**.980**
	1-2	.835	1-2	.837
PD	**4-5[a]**	**.945**	**4-6[b]**	**.882**
	1-2	.809	1-2	.794

Note. The maximum reliability estimate for each parameter is in boldface print. Below are the corresponding two-trial reliability coefficient estimates.

[a]The differences in variability explained by the daily schedule maximum reliabilities over that explained by the corresponding two-trial reliabilities were 35%, 25%, and 24%, for PP, AP, and PD, respectively. [b]The differences in variability explained by the twice-a-day schedule maximum reliabilities over that explained by the corresponding two-trial reliabilities were 16%, 26%, and 15%, for PP, AP, and PD, respectively.

daily schedule reliability for PP of .963 was not considerably different than .969 for the twice-a-day PP reliability (see Table 2). Although not considerably different, the daily schedule elicited maximum reliabilities with fewer trials administered. Namely, for the ideal daily schedule, four trials would be administered for both PP and AP, and five trials would be needed for the PD parameter. On the other hand, ideal reliability for the twice-a-day schedule would require six trials for each parameter (see Table 2).

For practical purposes, the reliability estimates for two-trial WAT administration, in which only the initial two trials were used for statistical analysis, have been included in Table 2. The practical reliabilities were considerably lower than the maximum reliability coefficient estimates (e.g., .963 for the PP daily maximum reliability compared to .762 for the two-trial schedule). The differences in variability explained for the maximum daily schedule over that explained by the corresponding two-trial reliabilities were 35%, 25%, and 24%, for PP, AP, and PD, respectively. As for the twice-a-day protocol, the differences in variability explained for the maximum twice-a-day schedule over that explained by the corresponding two-trial reliabilities were 16%, 26%, and 15%, for PP, AP, and PD, respectively (see Table 2).

Summary

The statistical analyses in this chapter demonstrated significant interactions for the PP and AP results between the daily and twice-a-day group and trial variables, indicating that the patterns of trial means for these results were not the same for both groups. The interaction was determined to be occurring because of significant differences between the trial-to-trial scores within each group rather than between the two groups at any specific trial. The significant differences within the groups trial-to-trial scores appears to occur primarily as a result of low initial trial scores when administering multiple-WAT schedules, and may indicate a learning effect.

Although not considerably different, the maximum reliability coefficient estimates can be determined with fewer trials administered with a daily testing schedule than with the twice-a-day schedule. Additionally, two-trial reliabilities, in which only the initial two trials were used for statistical analysis, explain far less variability than the maximum reliability schedules for both the daily and twice-a-day groups.

Chapter 5

DISCUSSION OF RESULTS

The purposes of this study were to investigate the trial-to-trial reliability of the WAT, determine whether any learning effect occurred with repeated WAT administration, and if so, to establish a trend-free trial schedule to derive reliable data. Intraclass reliability estimates were used to determine the trial-to-trial reliabilities of multiple-trial WATs. Two-way (Groups X Trials) ANOVAs for simple main effects were utilized to investigate trial-to-trial differences for the daily and twice-a-day schedules, within the groups and across the groups' trial scores. Additionally, a Tukey test was administered to identify where the trial differences were occurring. This chapter contains a discussion and interpretation of the results reported in Chapter 4. Sections in this chapter include the following: (a) trial-to-trial differences, (b) reliability estimates, (c) considerations for WAT protocol selection, and (d) summary.

Trial-to-trial Differences

In the present study, both testing schedules (i.e., the daily group and the twice-a-day group) yielded results in which PP and AP were lower in Trial

1 than in subsequent trials. This finding supported the primary research hypothesis, which proposed that a learning effect occurs in the initial two trials of a multiple-trial WAT. The learning effect was found to be complete by Trial 2, as substantiated by the lack of a significant change in the scores for Trials 2 through 6. These findings are in agreement with the report of Baumgartner and Jackson (1970), who stated that because subjects vary as to their readiness to perform at peak capacity (during long jump and physical perfomance tests), practice trials of the test are recommended. They also noted that it was often, and sometimes erroneously, assumed that subjects had previous knowledge of physical performance tests and therefore were usually tested without the benefit of practice. The nature of the WAT, being a supramaximal 30-s test, warrants consideration of a subject's previous experience with the procedures. In summary, preliminary practice trials seem necessary to eliminate any learning effect that may occur during the initial trial of a mutiple-trial WAT.

The same practice or learning effect in exerting maximal force on consecutive days or trials has previously been reported for static contractions (Laycoe & Marteniuk, 1971). Inbar and Bar-Or (1975), utilizing an anaerobic test similar to the WAT, also found that learning may be a factor, even after a 1-week separation between two consecutive testing sessions. Wilmore (1968), investigating work capacity tests, also identified a significant learning effect for work output and riding time between Day 1 and Day 2. Finally, Spodick (1975) submitted that investigators of physiologic and pathologic exercise responses have often determined that the first exercise session was a learning experience. All evidence makes it clear that it is very possible for a learning effect to occur in the initial two trials of a multiple-trial WAT. This conclusion leads to the recommendation that the prudent tester would always use practice trials when administering the WAT.

To further this argument for required practice, Dotan and Bar-Or (1983) suggested there is a need for muscular coordination or skill when performing repetitive, high-frequency motion, as is found in cycling during the WAT. Additionally, they stated that maximizing the short-duration power required for WAT performance was impeded by the circular motion of the pedals, which affected the nature of force application and increased the amount of skill and coordination neccessary for the given motion-sequence frequency.

Jacobs (1980) also found that the WAT demanded a high degree of neuromuscular coordination particularly during the high speed cycling in the initial seconds of the test. Therefore, initial practice trials would be critical to decrease the effect of these limiting factors in attempting to elicit maximal WAT performance. These previous findings support the results of the present study and provide further evidence suggesting preliminary learning trials will result in greater subsequent WAT performance.

Katch et al. (1974), using a modified anaerobic test, identified a small trial-to-trial and day-to-day practice effect occurring with multiple-trial testing. They proposed that preliminary practice trials be administered before any trial scores are used in subsequent data analysis. Additionally, Vandewalle et al. (1985), utilizing an anaerobic capacity test derived from the WAT, used the first two testing sessions for warm up and learning. Sharp et al. (1986), conducting a modified WAT of the upper body, initially gave the subjects a few minutes in order to practice the task before the actual testing was conducted. Finally, Katch et al. reported that if the absolute values of force output are to be compared between different samples of subjects, it would be crucial to establish that a plateau had been accomplished in the measured force-ouput scores. In the present study, a plateau for a trend-free trial schedule was established after the initial trial of the multiple-trial WAT, evidenced by the subsequent trial scores for both groups, namely, Trials 2 through 5 for PP in the daily group and Trials 2 through 6 for PP in the twice-a-day group, as well as for AP in both groups.

Interestingly, the results of this study revealed that there was no evidence of a learning effect for the PD parameter. Consequently, the fatigue rate was statistically the same for all trials in both groups. Therefore, if PD is the variable of primary concern during WAT testing, it may be unnecessary to use practice trials.

In summary, the results of this study substantiate the primary research hypothesis, which stated that a learning effect occurs during the initial trial of a multiple-trial WAT when the variables of interest are PP and AP. Additionally, based on the suggestions and results of past research, preliminary practice and/or learning trials should be administered before any data collection is attempted using the WAT. Finally, a plateau of trial scores is crucial in order to establish a trend-free trial schedule. A trend-free schedule was evident in the present study, indicating learning was no longer a factor in the derivation of data. As stated earlier, a plateau was found to occur after Trial 1 for PP and AP, therefore it would seem imperative to administer at least two trials in order to establish trend-free data.

Reliability Estimates

Although the daily and twice-a-day groups' maximum reliabilities were not considerably different, the daily schedule elicited slightly greater reliabilities with fewer trials administered. Therefore, due to one less trial to administer and elimination of an hour recovery period when using a twice-a-day schedule, daily schedule administration decreases the chance of a measurement error and the time it takes to administer a multiple-trial WAT.

With regard to reliability, WAT scores have been found to be well maintained in field situations where climate cannot be strictly controlled (Dotan & Bar-Or, 1980). A high reliability of performance in the WAT was reported for children, as well as adults (Bar-Or et al., 1977). Furthermore, Patton et al. (1985) determined the reliability for PP, AP, and PD for two resistance loads. Correlation coefficients for PP and

AP ranged from .91 to .93 for resistances of 4.41 and 5.59 J/rev/kg BW, respectively. The reliability coefficients for PD were considerably lower at .74 and .43 for both resistances. These values are consistent with the results of the present study, in which high reliabilities were seen for PP and AP, and the lowest reliability coefficients occurred with the PD parameter ($\underline{R}$ = .882) using the twice-a-day schedule.

In light of these high reliability estimates, it should be remembered that while the test-retest correlations are of considerable interest, they should be interpreted with caution (Wilmore, 1968). These coefficients are highly dependent on the size of the population and the degree of interindividual variability. A high coefficient would be expected when the sample size is small and the interindividual variability is high. In relation to the characteristics of the present study, the sample size was moderate ($\underline{N}$ = 25) and the interindividual variability was small, due to a rather homogeneous group of subjects. The majority of the subjects were college-aged athletes, either wrestlers or lacrosse players, which would lead to an expected similarity of scores. Interestingly, the maximum reliabilities in this study were rather high, ranging from .882 to .980, for the three anaerobic parameters tested. These reliabilities add to the already demonstrated high WAT reliability and are consistent with the previously published reliabilty coefficients (Bar-Or et al., 1977; Dotan & Bar-Or, 1980; Patton et al., 1985). The results from the present study further indicate that similar reliabilities are obtainable with the multiple-trial WAT protocols used in this study.

In summary, the daily schedule elicited slightly greater reliabilities with fewer trials administered than the twice-a-day schedule. Also, administering a multiple-trial WAT resulted in high reliabilty coefficients similar to those reported in previous WAT research.

Considerations for WAT Protocol Selection

It has been revealed through data analysis that a multiple-trial WAT test should be administered in order to maximize test scores, obtain acceptable reliability coefficients, and control a learning effect that occurs during the initial trial of WAT testing. Vandewalle et al. (1987) proposed that the choice of an appropriate anaerobic testing protocol should include only valid and reliable tests. The decision should also depend on the usefulness of the results, the aims of the testing, and the practicability of the tests within the testing sessions. The most reliable tests must be used if the purpose of the testing sessions is to compare individual results. The reliability of a test is less important if the average values of large populations have to be compared (Vandewalle et al.). Considering that WAT testing is often focused on a comparison of individual fitness results, utilizing the most reliable administration schedule would be justified.

In research, where precision of measurement is vital, determining a reliable criterion score that indicates maximum performance is neccessary, even if it requires administering many test trials (Baumgartner & Jackson, 1970). Coggan and Costill (1984), using 30-s and 60-s anaerobic tests, reported that more than four repeated trials would result in a better statistical estimate of a subject's true anaerobic performance. In the present study, two-trial reliabilities appear to be considerably lower than corresponding maximum reliabilities. The daily maximum reliabilities explained 35%, 25%, and 24% (for PP, AP, and PD, respectively) more variance than the corresponding reliabilities using only two trials. Also, the differences in variability explained for the twice-a-day maximum reliabilities over that explained by the corresponding two-trial reliabilities were greater by 16%, 26%, and 15% for PP, AP, and PD, respectively. Evans and Quinney (1981) suggested that accuracy within a test often has to be balanced against practical considerations (e.g., time to administer testing) when using a multiple-trial protocol. The decision of optimizing results by using four trials rather than two trials justifies the added expense of multiple-trial administration, however, this decision depends on the aims and time availability of the tester.

Dotan and Bar-Or (1983) revealed that a single test cannot optimize both PP and AP. Thus, optimal solutions may vary, depending upon whether the performance parameter of interest is PP, AP, or PD. The reliability coefficients and trial scores in the present study agree with this assessment, in that separate testing protocols were found to maximize test scores and reliability for PP, AP, and PD. Additionally, the daily schedule scores for PP and AP were 7% and 3% greater than the twice-a-day schedule scores, respectively, when comparing the corresponding individual trials across the groups. Therefore, based on the findings in the present study, in order to maximize reliabilities and optimize test scores, a daily four-trial WAT schedule should be used for PP and

AP. Trials 1 and 2 should be utilized for practice, then Trials 3 and 4 could be collected for results. As for PD, it is recommended that a daily five-trial WAT schedule be utilized, in which Trials 1 through 3 are used for practice, then Trials 4 and 5 could be collected for results with maximum reliability. As for which anaerobic parameter to investigate to best represent anaerobic performance, Patton and Duggan (1987) concluded that AP relative to BW is the best single index among those measured with the WAT.

In summary, based on previous WAT research and the present study, unless the WAT results are required within a 1- or 2-day period, it would be logical to administer a four- or five-trial WAT protocol using a daily schedule. For research purposes, in which maximum reliabilities and test scores are of utmost importance, it would be hasty and potentially inaccurate to base results on a two-trial WAT administration. There always exists the chance of a measurement, mechanical, and/or subject error at Trial 2, in which case the results, after eliminating the Trial 2 error, would be determined from Trial 1, which has been shown consistently to be a learning trial not representative of the subject's true maximum performance. Finally, it is recommended that scores and reliabilities will be optimized using a daily four-trial WAT schedule for PP and AP, whereas a daily five-trial WAT schedule should be used for PD.

<u>Summary</u>

The results presented in this study substantiate the primary research hypothesis, which stated that a learning effect would occur during the initial trial of a multiple-trial WAT. Additionally, preliminary practice and/or learning trials should be administered before any important data collection is completed. A plateau of trial scores is crucial in order to establish a trend-free trial schedule, in which learning is not a factor in the derivation of reliable data. In the present study, this was evident after the initial WAT trial.

The daily WAT testing schedule elicited maximum reliabilities with fewer trials to administer. The difference in variability explained for the daily and twice-a-day schedule maximum reliabilities over that explained by the corresponding values with only two trials administered was rather large. Whether the unexplained variability of two-trial testing is acceptable warrants consideration by each researcher. Additionally, administering multiple-trial WATs resulted in reliabilty coefficients similar to those reported in previous WAT research. Considering that the WAT is a supramaximal anaerobic test, results determined from a two-trial protocol are not advised. Therefore, for research purposes, in which maximum reliabilities and optimized test scores are of importance, it is suggested that a daily testing schedule using four or five WAT trials (depending upon the parameter of interest) be administered.

Chapter 6

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The purposes of this study were to investigate the trial-to-trial reliability of the WAT, determine whether any learning effect may occur, and if so, to establish a trend-free trial schedule to derive reliable data. This was accomplished by analyzing intraclass reliability coefficient estimates, as well as trial-to-trial scores in the daily and twice-a-day WAT testing schedules for each of three anaerobic parameters. The subjects were 25 males, ranging in age from 18 to 24 years. There were two experimental groups, consisting of a daily schedule (n = 12) and a twice-a-day schedule (n = 13). The daily subjects perfomed the WAT once a day for 6 consecutive days, whereas the twice-a-day group was administered the WAT twice a day for 3 consecutive days, with an hour recovery between trials on each day.

Trial-to-trial scores for both groups were analyzed with two-way (Groups X Trials) ANOVAs for simple main effects, as well as a Tukey test to identify where any significant differences occurred. PP, AP, and PD data were analyzed separately for each of the two groups. Intraclass reliability estimates were calculated to determine maximum reliabilities for the two testing protocols, as well as for a practical comparison with two-trial reliabilities.

Results revealed significant interactions between the daily and twice-a-day group and trial variables, indicating that the patterns for PP and AP were not the same for both groups. The interaction was determined to be occurring because of significant differences between the trial-to-trial scores within the groups rather than across the groups' trial scores. The significant differences within the groups' trial-to-trial scores appeared to occur primarily as a result of low initial trial scores when administering multiple-WAT schedules, and may indicate a learning effect which is complete by Trial 2.

Although the results are not considerably different, the maximum reliabilities for the daily testing schedule can be determined with fewer trials administered than with the twice-a-day schedule. Additionally, two-trial administration reliabilities explain far less variability than the corresponding daily and twice-a-day multiple testing maximum reliabilities. Also, the maximum reliabilities, for either schedule, have been found to be parameter dependent. Therefore, the WAT schedule the researcher selects depends on the anaerobic parameters of greatest interest, the time available for testing, and the degree of reliability and optimization of test scores that is acceptable.

Conclusions

The results of this study support the following conclusions:

1. There were significant differences within both groups in the trial-to-trial scores for PP and AP, but not for the PD parameter. The observed differences indicated that a learning effect was occurring between the initial two trials of WAT administration and this effect was found to be complete by Trial 2.

2. The statistical evidence indicated that the daily and twice-a-day schedules were not different for maximizing WAT scores. However, the PP and AP scores for the daily group were somewhat higher than the corresponding twice-a-day scores.

3. The practical administration of a simple two-trial WAT yielded reliabilities that explained far less variability than the corresponding daily and twice-a-day schedules. Therefore, if optimizing reliability is of concern, the tester is well advised to use at least a four-trial testing schedule.

4. It appears that multiple testing can increase the reliability of the WAT, and that a learning effect can be reduced by following a trend-free trial schedule as described in this study.

Recommendations

The findings of this investigation lead to these recommendations for further study:

1. A study should be conducted in which a 1-day three-trial testing schedule is utilized, to determine if optimal test scores and high reliabilities could be obtained in a 1-day period, which would economize the effort and time commitment of the subject.

2. A study should be conducted comparing the performance results of a WAT using the traditional resistance application and the instantaneous application of resistance.

Appendix A

INFORMED CONSENT FORM

1. a) Purpose of the study. To investigate the reliability of the Wingate Anaerobic Test in male subjects between the ages of 18 to 24 years.

b) Benefits. A major benefit to you will be a free evaluation of your current ability to perform high-intensity work. This may be of importance to you in evaluating your training status.

2. Method. You will be asked to report to the laboratory either three or six times, depending on the group you are placed in, in order to collect the appropriate data. The sessions will be spaced in order to minimize unwanted soreness on subsequent visits. Upon arriving at the laboratory your weight will be determined. You will be instructed as to the nature of the test, and then will be asked to warm up for 2 to 4 min. Following this there will be a 3- to 5-min rest interval. You will then be asked to maximally cycle for 30 s. You will be instructed to avoid pacing and maintain an all-out effort throughout the test. Following the testing session there will be a cool down, during which you will be encouraged to continue pedaling at a light workload for 2 or 3 min.

3. <u>Will this hurt</u>? All-out exercise is expected to be somewhat uncomfortable, but this will subside. Additionally, some muscle soreness may be experienced up to 48 hours following the exercise bout, but this is not expected to be severe. No lasting physical or psychological pain is expected from this study.

4. <u>Need more information</u>? Additional information may be obtained from Paul Schumann (273-3089) or Dr. Gary Sforzo (274-3359). All questions are welcomed and will be answered.

5. <u>Withdrawal from the study</u>. Participation is voluntary. You are free to withdraw from this study at any time without prejudice of any kind.

6. <u>Will the data be maintained in confidence</u>? Only the above mentioned researchers will have access to your data. Complete anonymity will be maintained, and once data is collected names will be replaced by codes.

7. I have read the above form and I understand its contents and I agree to participate in this study. I acknowledge that I am 18 years of age or older.

______________________________ ____________

Signature Date

Appendix B

MEDICAL HEALTH QUESTIONNAIRE

Name:

Date:

Age:

Sex:

Check if Yes

Past History

(In the past have you ever had?)

A. Rheumatic Fever ( )

B. Heart Murmur ( )

C. High Blood Pressure ( )

D. Heart Trouble:

Rhythm Abnormalities ( )

E. Disease of Arteries ( )

F. Varicose Veins ( )

G. Emotional Disorders ( )

H. Lung Disease ( )

I. Diabetes ( )

J. Epilepsy ( )

K. Injuries to back, knees, ankles, ( )

What kind?__________________

L. Stroke/Heart Attack ( )

How long ago?_______________

M. Operations ( )

What Kind?__________________

N. Other ( )

If other is checked explain here:__________________

Present Symptoms

(Have you recently had?)

A. Chest Pain ( )

B. Shortness of Breath ( )

C. Heart Palpitations ( )

D. Light Headedness ( )

E. Cough on Exertion ( )

F. Coughing up Blood ( )

G. Back Pain ( )

H. Arthritis ( )

I. Swollen Legs ( )

J. Chronic Thirst ( )

K. Loss of Consciousness ( )

Family History

(Have any blood relatives had?)

A. Heart Attacks ( ) Age ?

B. Heart Operations ( ) "

C. High Blood Pressure ( ) "

D. Diabetes ( ) "

E. High Cholesterol ( )

F. Congenital Heart Disease ()

G. Sudden Death () Age?

H. Other Major Illnesses ()

Explain Here:____________________

Risk Factors

	Yes	No
1. Smoking	()	()

How long?

How much?

2. Do you eat three meals a day Yes or No

Type of food mostly--red meat, fish, fruits & vegetables

3. Do you presently exercise three times a week for 20-30 min? Yes or No (circle response)

4. Is your occupation Sedentary (), Inactive (), Active (), Heavy Work () ?

5. Do you have discomfort, shortness of breath, or pain with exercise? Yes or No

6. Are you presently on medications? Yes or No

If yes, name the medications:____________________

Appendix C

WINGATE ANAEROBIC TEST DESCRIPTION

Now that your warm up has been completed, we are ready to start the actual testing. Your feet will be fastened to the pedals with toe clips, to the command "start," you will start pedaling to overcome the resistance of the bike. As soon as I set the resistance to the predetermined workload I will say "go," the test will begin and you will pedal as fast as you can for the next 30 s. You are instructed to avoid pacing and maintain an all-out maximal effort throughout these 30 s. Following this, there will be a brief cool down session. Do you clearly understand what you have to do to correctly perform this test?

REFERENCES

Allen, T. E., Smith, D., & Miller, D. K. (1977). Hemodynamic response to submaximal exercise after dehydration and rehydration in high school wrestlers. Medicine and Science in Sports, 9, 159-163.

Bell, D. G., & Jacobs, I. (1986). Electro-mechanical response time and rate of force development in males and females. Medicine and Science in Sports and Exercise, 18, 31-36.

Bar-Or, O., Dotan, R., & Inbar, O. (1977). A 30 second all-out ergometric test: Its reliability and validity for anaerobic capacity. Israel Journal of Medical Sciences, 13, 326-327.

Bar-Or, O., Dotan, R., Inbar, O., Rothstein, A., Karlsson, J., & Tesch, P. (1980). Anaerobic capacity and muscle fiber type distribution in man. International Journal of Sports Medicine, 1, 82-85.

Baumgartner, T. A., & Jackson, A. S. (1970). Measurement schedules for tests of motor performance. Research Quarterly, 41, 10-14.

Citterio, G., & Agostoni, E. (1984). Selective activation of quadriceps muscle fibers according to bicycling rate. Journal of Applied Physiology:

Respiratory, Environmental and Exercise Physiology, 57, 371-379.

Coggan, A. R., & Costill, D. L. (1984). Biological and technological variability of three anaerobic ergometer tests. International Journal of Sports Medicine, 5, 142-145.

Dotan, R., & Bar-Or, O. (1980). Climatic heat stress and performance in the Wingate Anaerobic Test. European Journal of Applied Physiology, 44, 237-243.

Dotan, R., & Bar-Or, O. (1983). Load optimization for the Wingate Anaerobic Test. European Journal of Applied Physiology, 51, 409-417.

Evans, J. A., & Quinney, H. A. (1981). Determination of resistance settings for anaerobic power testing. Canadian Journal of Applied Sport Science, 6, 53-56.

Frederick, F. A., Langevin, R. C., Milleti, J., Sacco, M., Murphy, M. M., & Patton, J. F. (1983). Development and assessment of the Monark cycle ergometer for anaerobic muscular exercise (Report No. T6/83). Natick, MA: United States Army Medical Research and Development Command.

Froese, E. A., & Houston, M. E. (1987). Performance during the Wingate Anaerobic Test and muscle morphology in males and females. International Journal of Sports Medicine, 8, 35-39.

Herbert, W. G., & Ribisl, P. M. (1972). Effects of dehydration upon physical working capacity of wrestlers under competitive conditions. Research Quarterly, 43, 416-422.

Inbar, O., & Bar-Or, O. (1975). The effects of intermittent warm-up on 7-9 year old boys. European Journal of Applied Physiology, 34, 81-89.

Inbar, O., & Bar-Or, O. (1986). Anaerobic characteristics in male children and adolescents. Medicine and Science in Sports and Exercise, 18, 264-269.

Inbar, O., Dotan, R., & Bar-Or, O. (1976). Aerobic and anaerobic components of a thirty-second supramaximal cycling task. Medicine and Science in Sports, 8, 51.

Inbar, O., Dotan, R., Trousil, T., & Dvir, Z. (1983). The effect of bicycle crank-length variation upon power performance. Ergonomics, 26, 1139-1146.

Inbar, O., Kaiser, P., & Tesch, P. (1981). Relationships between leg muscle fiber type distribution and leg exercise performance. International Journal of Sports Medicine, 2, 154-159.

Jacobs, I. (1980). The effects of thermal dehydration on performance of the Wingate Anaerobic Test. International Journal of Sports Medicine, 1, 21-24.

Jacobs, I., Tesch, P. A., Bar-Or, O., Karlsson, J., & Dotan, R. (1983). Lactate in human skeletal muscle after 10 and 30 s of supramaximal exercise. Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology, 55, 365-367.

Katch, F. I., McArdle, W. D., Pechar, G. S., & Perrine, J. J. (1974). Measuring leg force-output capacity with an isokinetic dynomometer-bicycle ergometer. Research Quarterly, 45, 86-91.

Katch, V. (1974). Body weight, leg volume, leg weight and leg density as determiners of short duration work, performance on the bicycle ergometer. Medicine and Science in Sports, 6, 267-270.

Kavanaugh, M. F., & Jacobs, I. (1988). Breath-by-breath oxygen consumption during performance of the Wingate test. Canadian Journal of Applied Sport Sciences, 13, 91-93.

Koutedakis, Y., & Sharp, N. C. C. (1986). A modified Wingate test for measuring anaerobic work of the upper body in junior rowers. British Journal of Sports Medicine, 20, 153-156.

Lamb, D. R. (1984). Physiology of exercise: Responses and adaptations (2nd ed.). New York: Macmillan.

LaVoie, N., Dallaire, J., Brayne, S., & Barret, D. (1984). Anaerobic testing using the Wingate and Evans-Quinney protocols with and without toe stirrups. Canadian Journal of Applied Sport Sciences, 9, 1-5.

Laycoe, R. R., & Marteniuk, R. G. (1971). Learning and tension as factors in static strength gains produced by static and eccentric training. Research Quarterly, 42, 299-306.

Mero, A. (1988). Blood lactate production and recovery from anaerobic exercise in trained and untrained boys. European Journal of Applied Physiology, 57, 660-666.

Palmer, W. K. (1968). Selected physiological responses of normal young men following dehydration and rehydration. Research Quarterly, 39, 1054-1059.

Patton, J. F., & Duggan, A. (1987). An evaluation of tests of anaerobic power. Aviation, Space, and Environmental Medicine, 58, 237-242.

Patton, J. F., Murphy, M. M., & Frederick, F. A. (1985). Maximal power outputs during the Wingate Anaerobic Test. _International Journal of Sports Medicine_, _6_, 82-85.

Safrit, M. J. (Ed.). (1976). _Reliability theory_. Washington: AAHPER Publications.

Saltin, B. (1964). Aerobic and anaerobic work capacity after dehydration. _Journal of Applied Physiology_, _19_, 1114-1118.

Sharp, N. C. C., Koutedakis, Y., Slater, T., Harris, G., & Katsikas, F. (1986). A test of anaerobic capacity of the upper body. _Scottish Journal of Physical Education_, _14_, 4-10.

Spodick, D. H. (1975). First and second versus second and third exercise tests. _Circulation_, _51_, 394.

Stevens, G. H. J., & Wilson, B. W. (1986). Aerobic contribution to the Wingate test. _Medicine and Science in Sports and Exercise_, _18_(Suppl. 6), 2.

Vandewalle, H., Peres, G., Heller, J., & Monod, H. (1985). All out anaerobic capacity tests on cycle ergometers. _European Journal of Applied Physiology_, _54_, 222-229.

Vandewalle, H., Peres, G., & Monod, H. (1987). Standard anaerobic exercise tests. _Sports Medicine_, _4_, 268-289.

Wilmore, J. H. (1968). Influence of motivation on physical work capacity and performance. Journal of Applied Physiology, 24, 459-463