

Criterion- Referenced and Norm- Referenced Agreement Between the Mile Run/Walk Test and PACER Test Among Male and Female Students of Sport Science Program

Assoc. Prof. Dr. Ahmad Hashim

Faculty of Sports Science and Coaching, Sultan Idris Education University, Malaysia

E-mail: ahmad@fsskj.upsi.edu.my

Dr. Mohd Sani Madon

Faculty of Sports Science and Coaching, Sultan Idris Education University, Malaysia

E-mail: mohd.sani@fsskj.upsi.edu.my

Sazali Hassan

Faculty of Sports Science and Coaching, Sultan Idris Education University, Malaysia

E-mail: sadzali@fsskj.upsi.edu.my

Yusop Ahmad

Faculty of Sports Science and Coaching, Sultan Idris Education University, Malaysia

E-mail: yusop@fsskj.upsi.edu.my

Abstract

The aim this study was conducted was to test the similarity between the reliability of the Mile Run/Walk Test and the PACER Test using criterion-reference and norm reference as a working framework. The sample for the study comprised 400 students, with male (N = 200) and female (N = 200) students randomly selected from the whole population students of Sports Science Program. Sample selection using the stratified random sampling selection strategy was administered using two attempts of the PACER test and one attempt of the Mile Run/Walk Test. The correlation between the test and repeat the test for the Test PACER male students is $r = .87$, $p < 0.05$. The correlation between test and re-test for the PACER test for female students was $r = .90$, $p < 0.05$. Altogether 83.5% male students and 80.0% females students passed the PACER Test based on the running time stipulated by the *FITNESSGRAM*® Standard (Cooper Institute for Aerobics Research, 2007). Altogether 63.5% male students and 68% female students passed the Mile Run/Walk Test based on the running time stipulated by the *FITNESSGRAM*® Standard (2007). The criterion-referenced reliability was estimated via the 'proportion of agreement' (Pa) and 'modified kappa' (kq) using the *FITNESSGRAM*® standard (2007). The similarity between the criterion-referenced reliability for the Mile Run/Walk Test and the PACER Test was average for male students ($Pa = .64$, $kq = .43$) as well as female students ($Pa = .65$, $kq = .38$),

Keywords: Reliability, Criterion-Reference and Norm-Reference, Mile Run/Walk Test and PACER Test

1. Introduction

For each student gaining maturity, an adequate aerobic ability level is normally linked with low level of heart disease, obesity, diabetes and hypertension. This problem is chronic and will usually occur during childhood. It is important to stress on education, achievement and the testing of aerobic ability from an early stage of physical development. It is clear that aerobic ability is important and is accepted as a component of students' fitness test. Physical Education and Sports Science Education teachers as well as school sports coaches can choose from several fitness programmes, including Presidents' Challenge (President's Council on Physical Fitness and Sports) or Prudential *FITNESSGRAM*® (2007) in order to assess aerobic achievement ability or cardiovascular. However, the *FITNESSGRAM*® fitness programme has been used comprehensively in national teenage fitness programmes and is used by the American Alliance for Health, Physical Education, Recreation and Dance (Safnt, 1995). All national teenage fitness programmes worldwide at present use criterion reference. This means that a special standard has been set for each test for each age group and sex. Subjects are then categorized as having passed if they fulfilled the standard and having failed if they do not achieve the required standard.

The criterion referenced standard for *FITNESSGRAM*® was developed to represent the fitness consistency level at the minimum level of adequate health and minimizes health risks (Morrow, Jackson, Disch & Mood, 1995). The component of the *FITNESSGRAM*® includes aerobic ability, body composition, muscle strength, muscle resistance and softness. The focus of this research was on the aerobic ability component. Physical Education and Sports Science Education teachers as well as sports coaches who use the *FITNESSGRAM*® have two choices to measure aerobic ability, namely the Mile Run/Walk Test and the PACER Test (Progressive Aerobic Cardiovascular Endurance Run) 20 meter multistage stage shuttle run. The development of the standard criterion reference for the Mile Run/Walk Test and the PACER are excellently explained in detail in an article by Cureton (1994).

The ability to track cardiovascular fitness (VO_{2peak} , $ml \cdot kg^{-1} \cdot min^{-1}$), and therefore, cardiovascular health status of large populations of sport science student using standard laboratory protocols (i.e., treadmill, metabolic cart) is not feasible. Thus, alternative valid and reliable field measures have been developed to indirectly assess cardiovascular health status. The 1-Mile Run/Walk and the Progressive Aerobic Cardiovascular Endurance Run (PACER or 20-meter shuttle run) have been established as the most valid and reliable field test when assessing the cardiovascular fitness (Beets & Pitetti, 2006). Undeniably, varied tests exist to measure cardiovascular fitness. Among them are the Mile Run/Walk Test, the PACER Test (20 Metre Multistage Shuttle Run), 1600 Meter Mile Run/Walk Test, $\frac{1}{4}$ Mile Run/Walk Test, $\frac{1}{2}$ Mile Run/Walk Test, $\frac{3}{4}$ Mile Run/Walk Test, 1.5 Mile Run/Walk Test, Nine- or 12-minute Run/Walk Test and other tests. However, none of the tests are consistent and used in all sport science program in our country.

Many Physical Education and Sports Science teachers as well as school sports coaches in our country have used the field test approach to measure students' cardiovascular fitness. In truth, we are merely measuring the cardiovascular fitness ability level based on the time taken using minutes and seconds to measure or in meter for the Mile Run/Walk Test, the 1.5 Mile Run/Walk Test, and the 12-minute Run/Walk. What is certain is that by using the testing approach, we are still unable to know students' real maximum volume of oxygen consumption. The Sports and Health Division, Education Ministry of Malaysia for example uses the 2,400 meter test in the Basic Physical Fitness Test as an instrument to measure forms one to five secondary school students' level of cardiovascular fitness. The *FITNESSGRAM*® includes both the Mile Run/Walk Test and PACER, leaving the decision on which to administer up to the teacher. And while the Mile Run/Walk Test and the PACER measure the same outcome (i.e., cardiovascular fitness), a question remains as to which assessment provides the most accurate and consistent results of current cardiovascular fitness across gender and age. That is, due to age, gender, or additional influences, certain populations of students may do better on one test when compared to the other (Beets & Pitetti, 2006).

1.1 Objectives of the Study

This study aimed to obtain the criterion referenced and norm referenced agreement between the Mile Run/Walk Test and the PACER Test among male and female students of Sport Science Program. Firstly, the findings of this research would determine the percentage of male and female students of Sport Science Program who passed the PACER test and the Mile Run/Walk Test based on the *FITNESSGRAM*® standard (2007). Secondly, the findings of this study would provide feedback on the reliability of the PACER test, which is the 20 meter multistage stage shuttle run. Thirdly, this was to determine the reliability of the Mile Run/Walk Test and the PACER Test so as to determine which test was the most suitable to measure the cardiovascular endurance among students depending on sex. Fourthly, the findings of this study would provide useful feedback to students of Sports Science Program, Faculty of Sports Science and Coaching Sultan Idris Education University on the appropriateness of tests to measure cardiovascular endurance among male and female students.

A study on the criterion referenced and norm referenced agreement between the Mile Run/Walk Test and the PACER Test (Progressive Aerobic Cardiovascular Endurance Run) among male and female students was conducted to achieve the following objectives. Firstly it was to test the test-retest reliability of the PACER test, the 20 meter multistage stage shuttle run. Secondly, it was to test the similarity in reliability of the Mile Run/Walk Test and the PACER Test based on the norm referenced and criterion referenced framework. Therefore, this was a study to find out if subjects could be classed as 'pass' or 'fail' when compared with the criterion reference for the Mile Run/Walk Test and the PACER Test using the standards set in the *FITNESSGRAM*®. Thirdly, this study was to determine the norm referenced reliability for the PACER and the correlation between the Mile Run/Walk Test in order to determine which test was the most suitable to be used in order to measure the cardiovascular endurance among students depending on sex.

2. Research Methodology

This section discusses the research methodology and the procedure for the Mile Run/Walk Test and the PACER Test among male and female students of Sport Science Program. The first section discusses the research sample, followed by the instruments and research procedure and also data analysis method. The study used the normative survey research approach. Data were analyzed with the aid of computer software. Descriptive statistical analysis was used to determine the level of maximum volume of oxygen consumption among male and female students. The Mile Run/Walk Test and PACER Test was completed based on criterion referenced and norm referenced of *FITNESSGRAM*® standard. Scores on level of maximum volume of oxygen consumption or cardiovascular fitness were obtained via the Mile Run/Walk Test and the PACER Test. Analysis was done using the Pearson Product Moment correlation approach in order to determine the reliability of the Mile Run/Walk Test and the PACER Test. The estimated criterion referenced reliability for the Mile Run/Walk Test and the PACER Test as well as were gained via the *proportion of agreement* (*Pa*) and modified kappa (*kq*).

2.1 Participants

The sample for the study comprised 400 students, with male (N = 200) and female (N = 200) students randomly selected from the whole population students of Sports Science Program, Faculty of Sports Science and Coaching Sultan Idris Education University. Sample selection using the stratified random sampling selection strategy was administered using two attempts of the PACER test and one attempt of the Mile Run/Walk Test. Approval for the study was obtained by the lecturer's and dean of Faculty of Sports Science and Coaching, because the tests were part of the regular sport science curriculum and were conducted on a routine basis, no informed consent was required from the students.

2.2 Instruments and Research Procedure

Progressive Aerobic Cardiovascular Endurance Run (PACER). The distance for the PACER was marked by painted stripes with red cones set at each end. Participants were instructed to run the distance between cones in the allotted time. The PACER protocol outlined by the *FITNESSGRAM*® Test User's Manual was used for all tests. The test was terminated either due to volitional exhaustion or because the participant could not keep up the required speed for two laps. The number of laps completed was recorded for data analysis by the sport science lecturer.

1-Mile Run/Walk. The 1-Mile Run/Walk Test was administered on an outdoor 400-meter athletic track (4 laps = 1 mile) according the *FITNESSGRAM*® testing procedures. Students were run in groups of no more than 12. One research assistant member timed the entire group using a digital hand-held stopwatch and read the elapsed time for each completed lap. The final fourth lap time (in minutes and seconds) was recorded by the sport science lecturer for each student.

2.3 Data Analysis

Criterion-Referenced Reliability and Agreement. Students were classified as being in either one of two conditions based on their 1-Mile Run/Walk Test and PACER performance for each trial. The criterion-referenced standards were based on sex specific performance standards outlined in the *FITNESSGRAM*® Test User's Manual. Criterion-referenced reliability was estimated using the proportion of agreement (*Pa*) and modified kappa (κ_q) statistics. For criterion-referenced reliability, the *Pa* is defined as the proportion of students classified the same on both trials of the PACER or 1-Mile Run/Walk Test. For criterion-referenced agreement, *Pa* is defined as the proportion of students classified the same on both the 1-Mile Run/Walk Test and PACER. To provide the most representative score, students' average 1-Mile Run/Walk Test and PACER performances for the two

trials were used to calculate the criterion-referenced agreement between the two tests. Means and standard deviations for all measures were calculated for male and female separately. Analyses Bivariate correlations were calculated to assess the relation between the students' average 1-Mile Run/Walk Test and PACER performance for the entire sample and for male and female separately. All analyses were conducted using Statistical Package for the Social Sciences (v 14.5).

3. Results

Table 1 shows the mean achievement for maximum oxygen consumption ($VO_2\text{max}$) in the PACER test based on sex of students. The mean achievement by male students was higher (Mean = 45.60 ± 3.86) $\text{ml.kg}^{-1}.\text{min}^{-1}$ compared with female students (Mean = 36.8 ± 3.27) $\text{ml.kg}^{-1}.\text{min}^{-1}$. based on performance standards outlined in the *FITNESSGRAM*® Test User's Manual.

Table 1: VO_2 Max Mean Achievement Based on the PACER Test According to Sex

Sex	Mean	SD
Male	45.60	3.86
Female	36.8	3.27

Table 2 shows the mean achievement for maximum oxygen intake ($VO_2\text{max}$) based on the Mile Run/Walk Test. The mean achievement by male students was higher (Mean = $8:39 \pm 1:67$) minutes compared with female students (Mean = 10:10, SD = 1:93) minutes based on performance standards outlined in the *FITNESSGRAM*® Test User's Manual.

Sex	Mean	SD
Male	8:39	1:67
Female	10:10	1:93

Table 2: VO_2 Max Mean Attained for the Mile Run/Walk Test (minutes:seconds) According to Sex

The statistical analysis indicated a positive and strong correlation between test scores and repeated tests for the PACER test on male and female students. The result showed that the correlation value between the test and the retest for the PACER Test on male students was value $r = 0.87$ and was significant at the $p < 0.05$ range. By comparison, the correlation value between the test and retest for the PACER Test on female students was value $r = 0.90$ and was significant at the $p < 0.05$ range. Statistical analysis using the Proportion of Agreement approach was used to determine the percentage of male and female students who could be classed as 'pass' or 'fail' compared with the criterion reference for the Mile Run/Walk Test using the *FITNESSGRAM*® standard reference (2007). Table 4 shows that only 63.5% of male students and 68.0% female students passed the Mile Run/Walk Test based on the running time specified by the *FITNESSGRAM*® Standard.

Table 4: Criterion Referenced Reliability of the Mile Run/Walk Test Compared with the *FITNESSGRAM*® Standard Based on Sex

Sex	Pass		Fail	
	freq	%	freq	%
Male	127	63.5	73	36.5
Female	136	68.0	64	32.0

Statistical analysis using the Proportion of Agreement approach was used to determine the percentage of male and female students who could be classed as ‘pass’ or ‘fail’ compared with the criterion reference for the PACER Test using the *FITNESSGRAM*® standard. Table 5 shows that 83.5% of male students and 80.0% of females students passed the PACER based on the running time specified by the *FITNESSGRAM*® Standard.

Table 5: Criterion Referenced Reliability of the PACER Test Compared with the *FITNESSGRAM*® Standard Based on Sex

Sex	Pass		Fail	
	freq	%	freq	%
Male	167	83.5	33	16.5
Female	160	85.0	40	15.0

Statistical analysis using the Proportion of Agreement and modified kappa was used to determine the relationship between tests among male and female students. Table 6 shows the relationship between the PACER Test and the Mile Run/Walk Test for male students based on the Proportion of Agreement value ($P_a = 0.64$) and Modified kappa ($k_q = 0.43$). The relationship between the PACER Test and the Mile Run/Walk Test for female students based on the Proportion of Agreement value ($P_a = 0.65$) and Modified kappa ($k_q = 0.38$).

Table 6: Similarity of Criterion Referenced Reliability between the PACER and The Mile Run/Walk Test

Test	Male	Female
PACER and the Mile Run/Walk	(n=200)	(n=200)
Proportion of Agreement	0.64	0.65
Modified Kappa	0.43	0.38

4. Conclusion and Discussion

Results of the study showed that the PACER Test had high reliability for testing the maximum volume of oxygen among male and female students. The findings of this study are similar with research findings by Liu, Plowman and Looney (1992) who reported multiple reliability for the

PACER Test via the test and retest approach, attaining .89 intra-class reliability value for 139 male and female subjects (Leger, Mercier, Gadoury & Lambert, 1998). As for the Mile Run/Walk Test, the findings show that only 63.5% of male students passed and succeeded in completing the run within a time range of 7:00 to 8:30 minute the time stipulated by the FITNESSGRAM® Standard. By contrast, the percentage of passes among female students in completing the run within a time range of 8:00 to 10:00 minute was 68.0%, the time determined by the FITNESSGRAM® Standard. As for the PACER Test, the results of the study showed that 83.5% male students passed and succeeded in completing the run between 72 to 106 repetitions based on the repeated runs stipulated by the FITNESSGRAM® Standard. Meanwhile, 85.0% female students passed and succeeded in completing the run between 41 to 72 repetitions. The result of the study showed that 16.5% male and 15.0% of female students failed to pass the level stipulated by the FITNESSGRAM® Standard.

The results of the study indicated a relation between the Mile Run/Walk Test and PACER Test for male students ($Pa = 0.64$) and ($Kq = 0.43$). There was also a relation between the Mile Run/Walk Test and PACER Test for female students ($Pa = 0.65$) and ($Kq = 0.38$). The amount of maximum oxygen use (VO_2^{\max}) is one of the best indicators to test aerobic fitness. However, subjects undergoing cardiovascular fitness tests need strong commitment to overcome various obstacles. Important aspects which are obstacles include instruments, a convenient place and subjects' ability as well as other aspects when undergoing cardiovascular fitness tests, especially one that involves measurement of maximum volume of oxygen consumption. Many researchers who have conducted such studies found several appropriate ways to measure the maximum volume of oxygen consumption. Among the ways are the Mile Run/Walk Test and the PACER Test which have not been tested comprehensively in Malaysia. The results of the study show that there is a relationship between the PACER Test and the Mile Run/ Walk Test for male and female students.

5. Recommendation

Based on the results of the study, it is recommended that within the state of Perak, attention needs to be given by all parties responsible for the physical development of Sport Science Program students' cardiovascular fitness. Essentially Physical Education and sport science teaching and learning activities be conducted efficiently by emphasizing the development of cardiovascular resistance. Similarly, activities for students such as sports curriculum and games should be given serious attention Physical Education and Sports Science school teachers ought to be more careful in selecting tests that are the most appropriate to measure cardiovascular resistance. It is recommended that teachers use the PACER Test to test students' cardiovascular resistance as it is proven to be appropriate and has very high test reliability compared with the Mile Run/Walk Test. It is suggested that the PACER Test can also be used as a component of the Physical Fitness Test battery to replace the Mile Run/Walk Test, which are still used by the Malaysian Ministry of Education, Teacher Education Division, Curriculum Development Centre and also State Education Departments to test students' cardiovascular resistance.

References

- [1] ACSM (1999). *ACSM's Fitness Book*. United State of America: Human Kinetics
- [2] Baumgartner, T.A, and Jackson, A.S., 1999, *Measurement for Evaluation (6th ed)*. USA: McGraw-Hill.
- [3] Baumgartner, T.A, Jackson, A.S., Mahar, M.T., & Rowe, D.A. (2003). *Measurement For Evaluation In Physical Education & Exercise Science (7th ed.)*. New York: McGraw Hill.
- [4] Beets, M.,W., & Pitetti, K., H. (2006). Criterion-Referenced Reliability and Equivalency Between the PACER and 1-Mile Run/Walk for High School Students. *Journal of Physical Activity & Health*, 3 (Suppl. 2), S21–S33
- [5] Bungum, T.J., Jackson A.W., & Weiller K.H. (1998). One-mile run performance and body mass index in Asian and Pacific Islander youth: Passing rates for the FITNESSGRAM®. *Research Quarterly for Exercise and Sport*, 69 (1), 89-94.
- [6] Carroll, B. (1994). *Assessment in Physical Education : A Teacher's Guide to the Issues*. London : The Palmer Press.
- [7] *Fitnessgram/Activitygram Test Administration Manual (4th ed)*. (2007). The Cooper Institute Dallas, Texas.
- [8] Cureton, K.J. (1994). Distance Running Performance Tests in Children: What do they mean? *Journal of Physical Education, Recreation and Dance*, 53(8), 64-66.
- [9] Franks, B.D., & Howley, E.T. (1998). *Fitness Leader's Handbook (2nd ed)*. USA : Human Kinetics.
- [10] Heyward, V.H. (1998). *Advanced Fitness Assessment in Exercise Prescription (3rd ed)*. USA: Human Kinetics.
- [11] Larsen, G.E., George, J.D., Alaxender, J.L., & Fellingham, G.W. (2002). Prediction of Maximum Oxygen Consumption from Walking, Jogging, or Running. *Research Quarterly for Exercise and Sport*, 73 (1), 66-67.
- [12] Leger, L.A., & Lambert, J. (1982). A Maximal Multistage 20-m Shuttle Run Test to Predict VO₂ max. *European Journal of Applied Physiology*, 49, 1-12.
- [13] Leger, L.A., Mercier, D., Gadoury, C., & Lambert, J. (1998). The Multistage 20 meter Shuttle Run Test for Aerobic Fitness. *Journal of Sport Sciences*, 6, 93-101.
- [14] Liu, N.Y.S., Plowman, S.A., & Looney, M.A. (1992). The Reliability and Validity of the 20 meter Shuttle Test in American Students 12 to 15 years old. *Research Quarterly for Exercise and Sport*, 63(4), 360-365.
- [15] Mahar, M.T., Rowe, D.A., Parker, C.R., Mahar F.J., Dawson, D.M., & Holt, J.E. (1997). Criterion-Referenced Agreement Between the Mile Run/Walk and PACER. *Measurement in Physical Education and Exercise Science*, 1 (4), 245-258.

- [16] McInnis, K.J., & Balady, G.J. (1999). Effects of Body Composition on Oxygen Uptake During Treadmill Exercise: Body Builders versus Weight-matched Men. *Research Quarterly for Exercise and Sport*, 70(2), 150-157.
- [17] Miller, D.K. (1994). *Measurement by the Physical Educator: Why and How*. USA: Human Kinetics.
- [18] Morrow, J. R., Jackson, A. W., Disch, J. G., & Mood, D. P. (1995). *Measurement and evaluation in human performance*. Champaign, IL: Human Kinetics.
- [19] Murray, TD., Walker, J.L., Jackson, A.S., Morrow, J.R., Eldridge, J.A., & Rainey, D.L. (1993). Validation of a 20-minute Steady State-jog as an Estimate of Peak Oxygen Uptake in Adolescents. *Research Quarterly for Exercise and Sport*, 64(1), 75-83.
- [20] Nieman, D.C. (2003). *Exercise Testing and Prescription. A Health-related Approach (5th ed)* New York: Mc-Graw-Hill.
- [21] Pate, R.R. (1991). Health Related Measures of Children's Physical Fitness. *Journal of School Health*, 61, 231-233.
- [22] Safrit, M.J. (1995). The validity and reliability of fitness tests for children. A review. *Pediatric Exercise Science*, 2, 9-28.
- [23] Vincent, S.D., Barker, R., Clarke, M., & Harrison, J. (1999). A Comparison of Peak Heart Rates Elicited by the 1-Mile Run/Walk and the Progressive Aerobic Cardiovascular Endurance Run. *Research Quarterly for Exercise and Sport*, 70(1), 75-79.