

The Validity and Reliability of Three Field Tests for Assessing College Freshmen Students' Cardiovascular Endurance

Heildenberg C. Dimarucot^{1,2,*}, Luisito S. Macapagal³

¹Department of Human Kinetics, College of Arts and Sciences, San Beda University, Manila, Philippines

²Department of Graduate Studies and Transnational Education, Institute of Education, Far Eastern University, Manila, Philippines

³Department of Mathematics and Statistics, College of Arts and Sciences, San Beda University, Manila, Philippines

Received December 29, 2020; Revised March 11, 2021; Accepted March 23, 2021

Cite This Paper in the following Citation Styles

(a): [1] Heildenberg C. Dimarucot, D.E.M, Luisito S. Macapagal , "The Validity and Reliability of Three Field Tests for Assessing College Freshmen Students' Cardiovascular Endurance," *International Journal of Human Movement and Sports Sciences*, Vol. 9, No. 2, pp. 363 - 374, 2021. DOI: 10.13189/saj.2021.090226.

(b): Heildenberg C. Dimarucot, D.E.M, Luisito S. Macapagal (2021). *The Validity and Reliability of Three Field Tests for Assessing College Freshmen Students' Cardiovascular Endurance*. *International Journal of Human Movement and Sports Sciences*, 9(2), 363 - 374. DOI: 10.13189/saj.2021.090226.

Copyright©2021 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract In an effort to reduce obesity rates among young adults resulting from their increasingly sedentary and stressful lifestyle and poor diet, three common endurance exercises that have been practiced in PE classes and also studied worldwide are *jumping rope*, *step test*, and *beep test* (20 meter multistage fitness test). This study shed light on the suitability of these three field tests as assessment tools for measuring the cardiovascular endurance of the college freshmen students of Physical Fitness (HKD-01) course since none has been investigated so far about this in the Philippine university context. Specifically, this study aimed to identify (1) the validity of the three tests as assessment tools for gauging cardiovascular fitness; and (2) their reliability as shown in the correlation between the directly measured and predicted VO₂ max. The statistical analyses revealed that the fitness tests were found valid and reliable. Each fitness test has its own peculiarity and course of action that make each distinct from one another. The step test is known to measure submaximal, the jumping rope test can be categorized as mid-maximal, and the beep test is known as a maximal test for cardiovascular endurance. It is then recommended that these fitness tests be employed in PE classes in the university following correct protocols to develop the cardiorespiratory endurance of the students.

Keywords Cardiovascular Endurance, Jumping Rope Test, Beep Test, Step Test, Field Test

1. Introduction

In recent years, university students have decreased physical activity participation and more and more of them develop sedentary behavior as they bridge to adulthood, enjoy autonomy, and make their personal lifestyle choices [1]. For instance, in one study in the US, it was reported that a small and declining minority of male and female college students are physically fit from 1996-2008. The male and female students' fitness levels gradually decreased over the past 13 years. These data are crucial especially since it was also implied in the study that physical activity and healthy nutrition can have a positive effect on the overall performance of the students [2].

Lang, Tremblay, Léger, Olds, and Tomkinson's review [3] revealed that the Philippines is in the 42nd percentile out of 50 countries in the 20m shuttle run performances of children and youth, indicating their low cardio-respiratory endurance. It was deemed as a public health issue in the country and in other developing countries. In another study, more than half of the Filipino students (67.2%) did not meet the recommended physical activity per week due to lack of time. The students do not sustain interest in physical fitness beyond their Physical Education classes due to more coursework as they progress in college and their poor

dietary intake. Even Physical Education majors themselves are not spared from this issue as they are also prone to living a sedentary lifestyle and possibly contracting chronic communicable diseases [4].

In fact, in one study in Brazil, it was found that 10% of the females and 7% of the males presented values of VO_2 max below the recommended levels for age, similar to other studies. Female students are expected to have lower cardiac output and oxygen transport capacity while males are physically more active than women--facts which some studies also supported [4]. Therefore, strengthening or revitalizing Physical Education (PE) classes and programs, particularly physical fitness courses, in the universities must be taken seriously.

To fulfill such demand, one of the objectives of the Physical Fitness Class (HKD-01) course of San Beda University is to assess the students' current physical fitness level through field tests to help the students achieve and ideally sustain a healthy lifestyle for lifelong learning. Starting the first semester, the teacher-researcher enhanced the course by adding jumping rope, step test, and beep test as fitness/field tests to measure the students' cardiovascular endurance. Previously, these tests were commonly used for varsity students' routine training, but no research has been done yet that checks the possibility of using the same tests for non-athletes, given the changing lifestyle and food intake due to technological advancements and fitness opportunities open for the new generation of students to seize more than ever.

One of the important indicators of physical fitness highlighted in the course is *cardiovascular (or cardiorespiratory, cardiopulmonary) endurance*. It "represents the combined ability of (1) the pulmonary system to exchange oxygen between the outside air and the blood circulating through the capillaries in the lung, (2) the cardiovascular system to transport oxygen to the working muscles, and (3) the muscular system to use oxygen." [5] Moreover, it also has mental health benefits along with other long-term effects like improved ability to extract oxygen from air during exercise, increased sweat rate, reduced risk of cancer, and increased density and breaking strength of bones, ligaments, and tendons, among many others. Thus, when one has an ideal cardiovascular endurance level, it would take less effort for that person to fulfill daily routine especially difficult tasks that require more physical exertion [6].

The best quantitative measure of cardiorespiratory endurance is maximal oxygen consumption or VO_2 max that is the amount of oxygen the body uses when a person reaches maximum ability to supply oxygen during exercise. However, this type of measurement is done in a laboratory and can be costly and time-consuming for a regular person [6]. Thus, some PE teachers only get an approximate measurement unlike what can be taken using a treadmill with an ECG machine in clinical practice.

Three common endurance exercises that have been practiced and also studied worldwide are *jumping rope*,

step test, and *beep test*.

In doing *jumping rope or skip test*, balance, and coordination abilities are crucial that it is considered as a skill (Trecroci, Cavaggioni, Caccia & Alberti, 2015) to sustain the precisely timed and rhythmic movements necessary for this exercise. It also develops the coordination of neuromuscular skills, muscle strength, and cardiovascular endurance [7]. Several studies, mostly from Asia, tested whether the inclusion of jumping rope routines or programs in PE classes would be beneficial to students--most of them with children as subjects [8, 9]. It was found that the training had significant effects on their cardiovascular endurance, flexibility, and muscular strength and endurance [10]. One recent study by Dimarucot and Soriano [5] employed the same fitness test and used the paired t-test to identify the significant difference in the VO_2 max of male and female university students after the jumping rope test. The study attested the effectiveness of the test in increasing their VO_2 max. Understandably, rope jump exercises could be found enjoyable for children to do and it is commonly practiced among athletes, but nothing much is known about its validity and reliability as a test to measure cardiovascular endurance of Filipino university students.

Step test is one of the popularly used assessments that provide reasonably close estimates of VO_2 max. It is considered a relatively quick and easy method for measuring cardiopulmonary fitness compared to other exercises or tests [11]. Knight, Stuckey, and Petrella [12] tested the validity of the prediction equation for VO_2 max used in the Step Test and Exercise Prescription (STEP) tool among a sample of 18 to 64 years old using the Bland-Altman analysis. There was a statistically acceptable agreement between the VO_2 max predicted using the STEP equation and directly measured through laboratory-based maximal treadmill testing. Cooney et al. [13] determined if the administration of the Siconolfi step test is a valid measure of cardiorespiratory fitness in patients with rheumatoid arthritis. Pearson correlation coefficient, intraclass correlation coefficient, and the Bland and Altman plots were used to determine the validity and reliability of the submaximal test. The Siconolfi step test provides a valid, reliable, and reproducible estimation of cardiorespiratory fitness (VO_2 max) in routine clinical practice, particularly in patients with rheumatoid arthritis. While the different varieties of step test were studied and tried clinically for people with special conditions and non-clinically for college students [14, 11] no study has been conducted on the STEP tool by Knight, Stuckey, and Petrella [12] as tested and applied in the Philippine context and no comparison on the performance of young non-athletes using the same test and statistical analysis was done so far.

Lastly, the *beep test* (also called multistage fitness test [MSFT] or shuttle run test [SRT]) has been widely used as a predictive test for VO_2 max, originally designed to predict fitness in healthy adults attending fitness classes

and in athletes participating in sports. It is used by at least 50 countries due to the low cost of equipment, simplicity in administering and scoring the test, flexibility in testing, convenience for the participants, and practical use for simultaneously testing large groups of children and youth [3]. Paradisis et al. [15] investigated the validity and suitability of predicting VO_2 max and velocity VO_2 max of PE students on the basis of their performance in the 20m multistage fitness test (MSFT) using Pearson's product-moment correlation coefficient. Results indicated a high correlation coefficient between shuttles in the 20m MSFT and VO_2 max as well as vVO_2 max. It was then concluded that the 20m test can accurately predict VO_2 max and vVO_2 max and this field test can provide useful information regarding the aerobic fitness of young adults. On the other hand, Mayorga-Vega, Aguilar-Soto, and Viciano [16] used the Hunter-Schmidt's psychometric meta-analysis approach in their own study and it was revealed that the test has a moderate-to-high mean correlation coefficient of criterion-related validity for estimating VO_2 max and that its validity is higher for adults than for children. It was emphasized, however, that test scores are just an estimation of cardiorespiratory fitness.

Though the beep test was found to be an effective and accurate international population health indicator in children and youth [17, 18, 19, 20, 21, 22, 23], it is still worth examining whether the results are particularly true as well in the Philippine context, especially when this test is compared with the first two field tests above using a different statistical analysis tool.

Thus, this present study shed light on the suitability of the three field tests as assessment tools for measuring the cardiovascular endurance of the college freshmen students of the Physical Fitness Class (HKD-01) course at San Beda University. Specifically, this study aims to identify (1) the validity of the three tests as assessment tools for gauging cardiovascular fitness and (2) the reliability as revealed by the correlation between the directly measured and predicted VO_2 max.

2. Materials and Methods

2.1. Research Design

A cross-sectional study was conducted that involves examining the subjects who differ in cardiovascular endurance level at one specific point in time. The data are collected at the same time from the participants who are similar in age but different in anthropometric factors such as body mass index (BMI), waist circumference, waist to hip ratio, body fat percentage, and heart rate.

2.2. Research Participants

All the anthropometric measurements in the three field

tests were taken from all the HKD-01 (Physical Fitness Class) students of San Beda University who voluntarily agreed to participate in the study. The study included apparently healthy male and female students whose ages range from 18 to 20 years who are also non-smokers and non-alcoholics. The latest version of the Physical Activity Readiness Questionnaire known as (PAR-Q + 2019) as a screening test for the participants was also used to identify those who are qualified to undergo the test. The study did not cover those students who have a history of any acute or chronic illness, were on medication and were undergoing regular physical training (as an athlete or non-athlete). Those who answered one YES or more in any of the questions in the 2019 PAR-Q were automatically disallowed to participate in the study.

2.3. Measurement and Instrumentation

The anthropometric variables like height and weight, heart rate, waist circumference, and waist to hip ratio were measured using a tape measure, a digital weighing scale, and a pulse oximeter. The waist-to-hip ratio was computed by dividing the waist circumference with the hip measurement. The formula for getting the body mass index (BMI) is kg/m^2 where kg is a person's weight in kilograms and m^2 is their height in meters squared [24]. For the body fat percentage, the prediction equation of [25] was used which was also adopted by [9]:

$$\text{Percentage body fat} = 0.610 \times (\text{sum of skinfold of triceps and calf}) + 5.$$

The formula that was used to assess the cardiovascular endurance levels of the students [24] is as follows:

$$\text{VO}_2 \text{ max} = 132.6 - (0.17 \times \text{weight in kg}) - (0.39 \times \text{age}) + (6.31 \times \text{gender [0 for females; 1 for males]}) - (3.2 \times \text{time to walk 1 mile}) - (0.156 \times \text{post-exercise heart rate [bpm]}).$$

2.4. Data Collection Procedures

The following protocols were observed for each of the three field tests:

2.4.1. Jumping Rope Test

Table 1 shows the rope jumping test routine for eight weeks held at the school gym. This was slightly patterned after Partavi's rope-jump training program [8]. It starts with the usual warm-up exercise which is dynamic stretching for 10 minutes except for the first week with only 5 minutes. Since the jumping rope test that was done is a progressive and multistage one, 100 repetitions were added per week which starts at 100 in Week 1 and ends at 800 in Week 8. All tests end with a 5-minute static stretching for the cool down. The anthropometric measurements as well as the VO_2 max are calculated in the pre-test and the post-test.

Table 1. Jumping Rope Test Eight-Week Routine

Week	Intensity (jumps/min)	Warm up: Dynamic Stretching	Exercise	Cool down: Static Stretching
1	100	5 mins.	100 repetitions per minute with 30-second rest	5 mins.
2	200	10 mins.	200 repetitions per minute with 30-second rest	5 mins.
3	300	10 mins.	300 repetitions per minute with 30-second rest	5 mins.
4	400	10 mins.	400 repetitions per minute with 30-second rest	5 mins.
5	500	10 mins.	500 repetitions per minute with 30-second rest	5 mins.
6	600	10 mins.	600 repetitions per minute with 30-second rest	5 mins.
7	700	10 mins.	700 repetitions per minute with 30-second rest	5 mins.
8	800	10 mins.	800 repetitions per minute with 30-second rest	5 mins.

2.4.2. Step Test

The Step Test and Exercise Prescription (STEP) tool developed by Knight, Stuckey, and Petrella [12] was the basis for the protocol in this study. The test begins with a demonstration of the stepping pattern to the students: stepping one foot at a time up both steps and back down again. One step is counted after stepping up to the top step and returning back down to the starting position. When the student is ready to begin, the teacher starts the timer as the first foot leaves the ground. The teacher has to offer encouragement and watch for signs of fatigue or balance problems during the test. As the student performs the last step, the teacher takes the heart rate, palpating radially in a six-second count. Next, the teacher calculates the patient’s target training heart rate, using 65% to 85% of the age-predicted maximal heart rate as a target.

2.4.3. Beep Test (20 Meter Multistage Fitness Test)

The protocol is a slight modification of the same test protocol implemented by Voss and Sandercock [26]. The students ran 20-meter shuttles in time to an audible metronome at an initial speed of 8.5 km h⁻¹ increasing by 0.5 km h⁻¹ each minute. All testing was inside the school gym and not in an open area or field unlike in other studies. The testing was conducted in groups of up to 30 with a ratio of five participants to one teacher-researcher. Participants were not practiced in this test prior to the pre-test, but they performed it for 8 weeks as one of their lessons on cardiovascular endurance. The teacher-researcher reminded the subjects to run for as long and as best as possible as this test will predict their present cardiovascular endurance level. The teacher also acted as a spotter during the test and recorded the participants’ final shuttle count at either the point of volitional exhaustion or when they have failed thrice. The number of completed shuttles was then converted to running speed in km/h⁻¹ at the final completed stage. The participants had their training heart rate with an eight-second interval within one minute then they rested for 1 minute. Lastly, they got their recovery heart rate for another minute which was used for the VO₂ max computation later. After each field test, both pre- and post-test, the VO₂ max were computed and subjected to statistical analysis.

2.5. Statistical Tools

A paired-samples t-test was conducted to determine if there are significant differences between the VO₂ max measured using the different fitness tests and measured using a portable machine. This test was utilized to assess the validity of the different fitness tests in measuring VO₂ max.

To determine the reliability of the different fitness tests, several tests were used. First, the Bland-Altman plot (limits of agreement or different plots) and analysis is a method comparison technique [27] used for assessing agreement and average bias (or difference) between two measurements techniques and assumes that data are paired. As suggested by Twomey [27], the relative difference plot [28] was used in this study for two reasons: 1) the presence of ‘funnel’ effect in the absolute difference plot, that is, there is an increase in the degree of differences as mean values increase, 2) by inspection, the slopes in the data are significantly different from zero, and 3) the histogram and normality tests indicated the non-normality of the differences. Using this plot and analysis, two methods are considered to be in agreement and may be used interchangeably if the limits of agreement (or LoA, the mean difference ± SD of differences) do not exceed the maximum allowed difference Δ [29]. Further, the methods do not disagree if the maximum allowed difference Δ is higher than the upper 95% CI limit of the higher LoA and -Δ is lower than the lower 95% CI limit of the lower LoA, as shown in Figure 1.

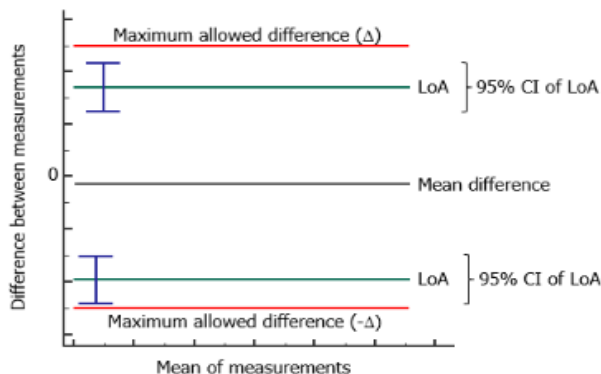


Figure 1. Bland-Altman Plot (2020) from the MEDCALC website (<https://www.medcalc.org/manual/blandaltman.php>)

Another statistical test used is the one-sample t-test for a mean. The Bland–Altman plot allows identification of any systematic difference between the measurements (i.e., fixed bias). The mean difference is the estimated bias, and the standard deviation of the differences measures the random fluctuations around this mean. The presence of fixed bias is evident when the mean value of the difference differs significantly from 0 on the basis of a one-sample t-test [30].

This study also utilized tests of significance of the standardized coefficients resulting from simple linear regression analysis in order to detect the presence of proportional bias (the percentage difference between the measures is a function of the average of the measures) than those from the other by an amount that is proportional to the level of the measured variable). Additionally, the intra-class correlation coefficient was also used in performing reliability testing. It is an improvement compared to Pearson's (as what was used in Paradisis et al. [15]) and Spearman's as it takes into account the difference in the measures for individual subjects, along with the correlation between methods.

3. Results

3.1. Characteristics of the Sample

A total of 115 freshman students of the San Beda University participated in this study, 52.2% of which are male. Table 2 presents the characteristics of the sample.

As presented in Table 2, the average age of the sample was 18.29 ± 0.542 years, the average height was 160.79 ± 14.083 cm, the average weight was 77.784 ± 38.379 kgs. The average BMI was 23.213 ± 8.523 kg/m² which is under the “at-risk” level based on the standards posted on the Philippine Association for the Study of Overweight and Obesity website. The average resting heart rate was 69.69 ± 16.938 bpm which is within the normal range for adults based on the Cleveland Clinic website.

Table 2. Sample Characteristics

	Mean	Std. Dev.	Lowest – Highest
Age (years)	18.29	.542	17 – 20
Height (cm)	160.79	14.083	60 – 187
Weight (kg)	77.784	39.879	37 – 262.3
Waist to Hip	.850	.162	.68 – 1.80
Body mass index (kg/m ²)	23.213	8.524	12.8 – 96.1
Fat	24.161	7.284	6.4 – 38.7
Resting heart rate (bpm)	69.790	16.948	18 – 147

3.2. Validity of the Fitness Tests: Tests of Difference between Predicted and Directly Measured VO₂max

To determine whether there are significant differences that exist between the predicted (Test_1) and directly measured (Test_2) VO₂ max in the three different fitness exercises, the t-test for dependent groups was used.

The results shown in Table 3 indicate that in the beep test, there is a significant difference between the predicted VO₂ max ($M = 30.25, SD = 4.710$) and the directly-measured VO₂ max ($M = 31.80, SD = 5.686$) since $t(114) = -5.728, p < .01$. In fact, it can be concluded that the mean predicted VO₂ max is significantly lower than the mean directly-measured VO₂ max mean difference (-1.542).

On the other hand, results do not support significant differences between the mean predicted and mean directly-measured VO₂ max when the subjects underwent the jumping rope test, $t(114) = -.7697, p > .05$, and the step test $t(114) = 2.4963, p > .05$.

3.3. Reliability of the Three Fitness Tests: Bland-Altman Plot and Analysis

Figures 2 to 5 present the Bland-Altman plots that were generated to assess agreement between the predicted VO₂ max arising from the different fitness tests and the directly-measured VO₂ max.

Table 3. Summary of Descriptive Measures

VO ₂ max	Lowest – Highest	Mean	Std. Dev.	Skewness	
				Statistic	Std. Error
Predicted_skip	18.93 – 17.01	39.40	10.733	.981	.226
Measured_skip	10.53 – 68.49	40.17	8.976	.513	.226
Predicted_beep	26.80 – 43.90	30.25	4.710	1.381	.226
Measured_beep	26.80 – 50.80	31.80	5.686	1.209	.226
Predicted_step	18.93 – 77.73	43.83	14.470	.951	.226
Measured_step	18.93 – 71.07	41.33	9.429	1.021	.226

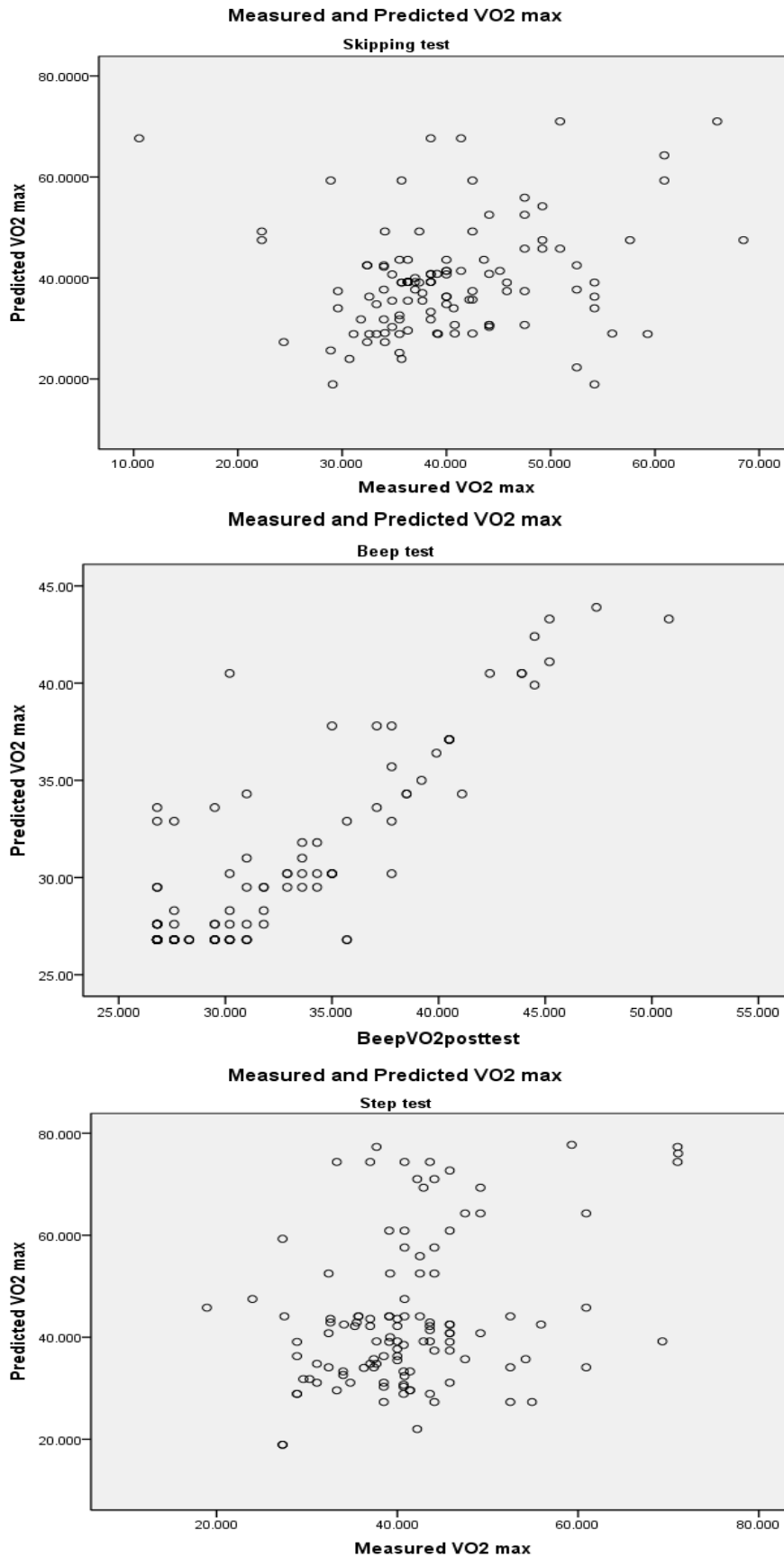
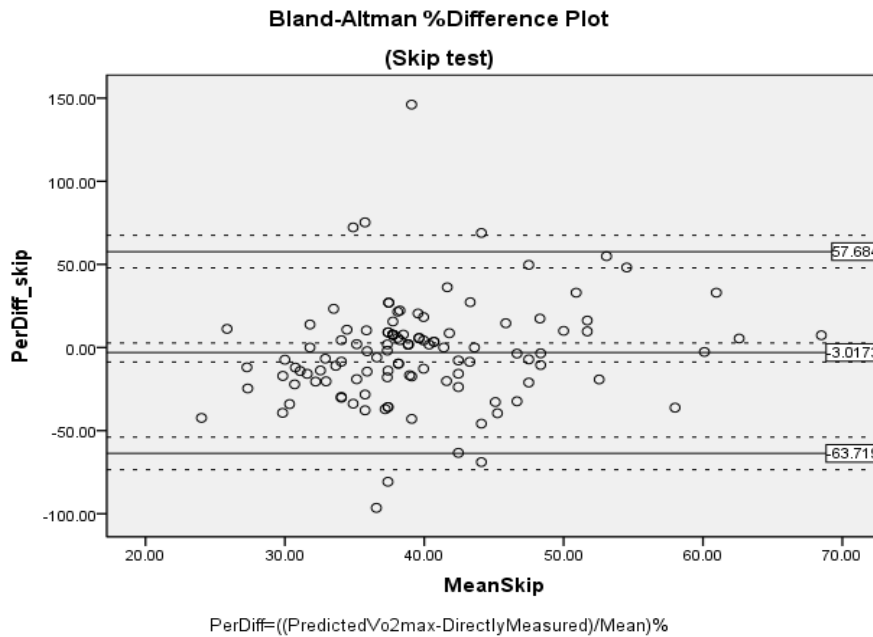


Figure 2. Scatterplot of Measured and Predicted VO₂ max of the Three Fitness Tests

Table 4. Results of Paired Samples T-Test of the Three Fitness Tests

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Dev.	S.E. Mean	95% CI of the Difference				
				Lower	Upper			
Skip_1 – Skip_2	-0.770	12.556	1.171	-3.089	1.550	-0.657	114	.512
Beep_1 – Beep_2	-1.542	2.886	.269	-2.075	-1.009	-5.728*	114	.000
Step_1 – Step_2	2.496	14.359	1.339	-.156	5.149	1.864	114	.065

*difference is significant at $p < .05$

**Figure 3.** Plot of Differences between Percentage of Differences and Mean of the Two Measurements (Skip Test and Directly-Measured VO₂ max)**Table 5.** Bland-Altman Analysis: Bias and Limits of Agreement for Predicted VO₂ max (Skip Test) Versus Directly-Measured VO₂ max

Parameter	Count	Value	Standard Deviation	Standard Error	95% confidence interval	
					lower	upper
Difference	115	-0.7698	12.556	1.171	-3.089	1.550
%difference	115	-3.0173	30.970	2.888	-8.738	2.704
Lower LoA	115	-63.719		4.938	-73.524	-53.913
Upper LoA	115	57.684		4.938	47.878	67.490

Results shown in Table 4 indicate that in the Beep test, there is a significant difference between the predicted VO₂ max () and the directly-measured VO₂ max () since. In fact, it can be concluded that the mean predicted VO₂ max is significantly lower than the mean directly-measured VO₂ max, mean difference = -1.542.

On the other hand, results do not support significant differences between the mean predicted and mean directly-measured VO₂ max when the subjects underwent the skipping test, and the step test.

Figure 3 shows the Bland-Altman plot of differences between the % difference and the mean of the predicted VO₂ max (skip test) and directly-measured VO₂ max. Here, it is observed that there is a very slight underestimation

since the mean percentage difference between the predicted and the directly-measured VO₂ max is very slightly below 0 (mean percentage difference = -3.017%, with 95% confidence interval of -8.738% to 2.704%). Thus, the predicted VO₂ max obtained after the respondents performed the skip test underestimated the directly-measured VO₂ max by an average of 3.017%. The variability of the percentage differences appears to be +/- 60%, which is somewhat wide. There seems to be some random variation of the percentage difference around the constant mean.

Table 5 presents the bias (% difference) and limits of agreement together with their corresponding 95% confidence interval (CI) estimates. These are how precise

the ballpark figure is. The 95% CI of the mean % difference illustrates the magnitude of the systematic difference [31]. In this case, there is no significant systematic difference or fixed bias since the line of equality (horizontal line through 0) is in the interval (-8.738 to 2.704). The presence of fixed bias is evident when the mean value of the difference differs significantly from 0 on the basis of a one-sample t-test, with results shown in Table 6 $t(114) = -1.045, p > .05$. The 95% CI of the level of agreement (LoA) limits allows for the estimates of the size of the possible sampling error (Giavarina, 2015). It can be observed that the CIs, both the mean % difference (+/-5.72%) and the LoA limits are narrow (+/-9.81%) because of the considerably large sample size used in this study.

Figure 4 shows the Bland-Altman plot of differences between the percentage difference and the mean of the predicted and directly-measured VO₂ max after the beep test. Here, it can be observed a very slight underestimation since the mean percentage difference between the predicted and the directly-measured VO₂ max is slightly below 0 (mean percentage difference = -4.609%, with a 95% confidence interval of -6.236% to -2.982%). Thus, the predicted VO₂ max obtained underestimated the directly-measured VO₂ max by an average of 4.609%. The variability of the percentage differences appears to be +/- 17%, which is narrow. However, the variability around the mean does not appear to be constant. There seems to be an observable downward trend when the mean VO₂ max is greater than 40.

Table 6. One-Sample Test of Difference

	Test Value = 0					
	T	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
%Diff_skip	-1.045	114	.298	-3.01735	-8.7384	2.7037
%Diff_Beep	-5.612	114	.000	-4.60899	-6.2360	-2.9820
%Diff_Step	1.084	114	.281	3.18677	-2.6391	9.0127

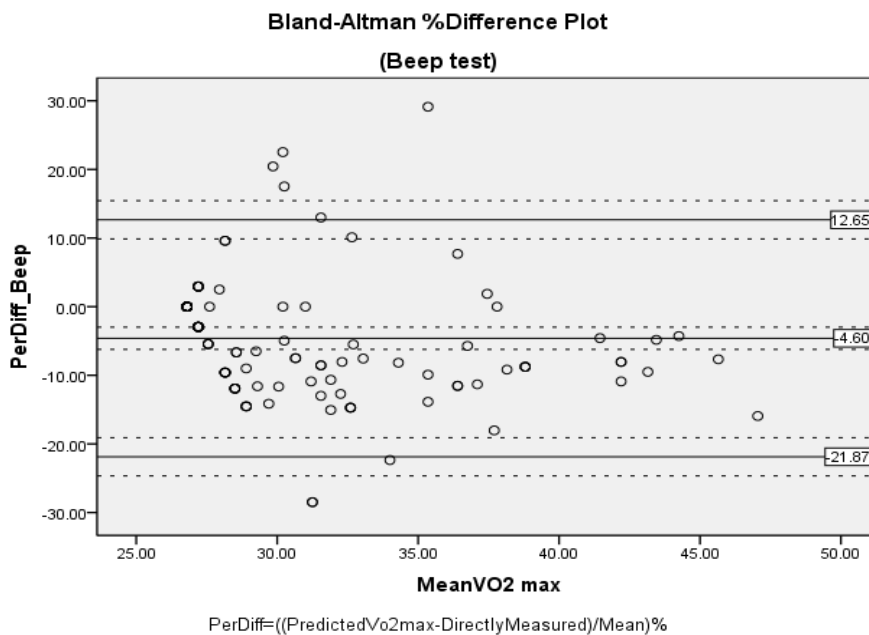


Figure 4. Plot of Differences between Percentage of Differences and Mean of the Two Measurements (Beep Test and Directly-Measured VO₂ max)

Table 7. Bland-Altman Analysis: Bias and Limits of Agreement for Predicted VO₂ max (Beep Test) Versus Directly-Measured VO₂ max

Parameter	Count	Value	Standard Deviation	Standard Error	95% confidence interval	
					lower	upper
Difference	115	-1.542	2.886	0.269	-2.075	-1.009
%difference	115	-4.609	8.808	0.821	-6.236	-2.982
Lower LoA	115	-21.873		1.405	-24.661	-19.084
Upper LoA	115	12.655		1.405	9.866	15.443

Table 7 presents the bias (% difference) and LoA together with their corresponding 95% confidence interval (CI) estimates. In this case, there is a significant systematic difference since the line of equality (horizontal line through 0) is not in the interval (-6.236 to -2.982). This implies that the predicted VO₂ max obtained after the respondents undergo the Beep Test more consistently underestimates the directly-measured VO₂ max. The presence of fixed bias is supported by the results of the one-sample t-test, as shown in Table 6, $t(114) = -5.612, p < .05$. It can be observed that the CIs, both the mean percentage difference (+/-1.63%) and the LoA limits (+/-2.79%) are narrower.

Figure 5 shows the Bland-Altman plot of differences between the percentage difference and the mean of the predicted VO₂ max (step test) and directly-measured VO₂ max. Here, a very slight over estimations since the mean percentage difference between the predicted and the

directly-measured VO₂ max is slightly above 0 (mean percentage difference = 3.1867%, with a 95% confidence interval of -2.639% to 9.013%). Thus, the predicted VO₂ max underestimated the directly-measured VO₂ max by an average of 3.1867%. The variability of the percentage differences appears to be +/- 62%, which is somewhat wide. The variability around the mean does not appear to be constant. Noticeably, there seem to be more positive percentage differences when the mean VO₂ max is greater than 55.

Table 8 presents the bias (% difference) and LoA together with their corresponding 95% confidence interval (CI) estimates. In this case, there is no significant systematic difference since the line of equality (horizontal line through 0) is in the interval (-2.639 to 9.013). It can be observed that the CIs, both the mean % difference (+/-5.83%) and the LoA limits (+/-9.99%) are also narrow

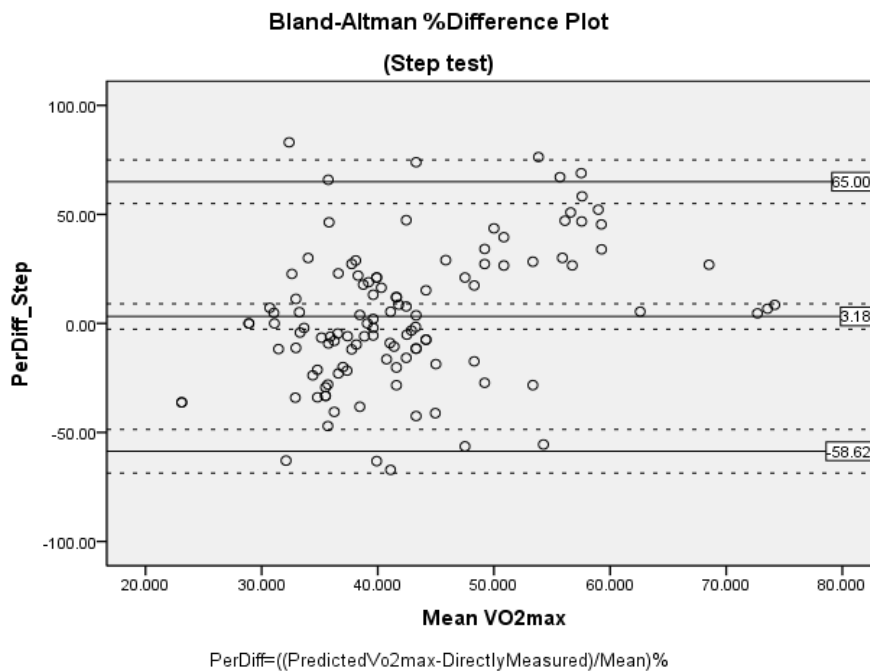


Figure 5. Plot of Differences between Percentage of Differences and Mean of the Two Measurements (Step Test and Directly-Measured VO₂ max)

Table 8. Bland-Altman Analysis: Bias and Limits of Agreement for Predicted VO₂ max (Step Test) Versus Directly-Measured VO₂ max

Parameter	Count	Value	Standard Deviation	Standard Error	95% confidence interval	
					lower	upper
Difference	115	2.496	12.359	1.339	-0.156	5.149
%difference	115	3.187	31.538	2.941	-2.639	9.013
Lower LoA	115	-58.627		5.029	-68.613	-48.642
Upper LoA	115	65.001		5.029	55.016	74.987

Table 9. Results for Regression of Mean % Difference

	Predicted VO ₂ max Skip test	Predicted VO ₂ max Beep test	Predicted VO ₂ max Step test
Standardized coefficients	.183	-.255	.369
t-values	1.983	-2.806**	4.219**

**significant at $p < .05$

3.4. Significance of Standardized Coefficients from Simple Linear Regression Analysis

Table 9 shows the standardized coefficients resulting from the simple linear regression analysis. Notably, the proportional bias is present in the cases of predicted VO₂ max of the beep test ($= -.255, p < .05$) and predicted VO₂max of the step test ($= .369, p < .05$).

4. Discussion

There are no significant differences between the mean predicted and the mean directly-measured VO₂ max when the subjects underwent the skipping test and the step test, except for the beep test. The predicted VO₂ max obtained after the respondents underwent the three tests underestimated the directly-measured VO₂ max, with wide variability except for the results in the beep test. There is no significant systematic difference in the skip and step test, but not the same case in the beep test. There seems to be a random variation of the percentage difference around the constant mean from the results for the three tests.

Thus, despite the three tests' apparent differences, it can still be concluded that the tests are valid and reliable tools to be used for university non-athlete students, similar to the results of the study of Knight, Stuckey, and Petrella [12] and Cooney et al. [13] for step test that also used Bland-Altman analysis; and Paradisis et al. [15] and Mayorga-Vega, Aguilar-Soto, and Viciano [16] for beep test, though the study used different statistical tools. This was, however, the first study so far that has tested the validity and reliability of the jumping rope test for university students that can be used for comparison in future studies.

The Bland-Altman method defines the intervals of agreements, but not the limits of acceptability. Thus, it can be noted that though the tests have been found suitable and practical assessment tools based on the analyses, they have different levels of complexity that contributed to their differences. These tests then need to be used considering the purpose of the use and the physical readiness of the students.

The step test is a procedure used to assess submaximal oxygen uptake, best for beginners. Its procedure is easy to follow. This does not pose any restrictions on the participants because it is not as strenuous as an assessment compared to the other two. It does not need any inclusion and exclusion criteria. Through this test, it is easy to

measure the approximate cardiovascular endurance provided it is done with less space, used with a metronome, and executed following a strict protocol.

The jumping rope test, on the other hand, is known as a skill-based fitness test that is ideal for intermediate level due to its skill familiarization and acquisition. The participants must have sufficient time to practice the skill that includes mastering the right timing and coordination. It cannot be classified as an assessment for submaximal because it requires higher skills that have to be acquired through constant practice. This assessment was not commonly used with university students because it is somewhat tedious as teachers need to teach the additional skills mentioned and they also need to motivate students to do it by making it a fun activity aside from acquiring the skill themselves.

Lastly, the beep test is for advanced level as it requires rigid screening assessment and strict compliance of procedure. This explains its notable difference from the jumping rope test and the step test in the analyses. It gives an accurate assessment of cardiovascular endurance, but this requires inclusion and exclusion criteria because it uses maximal running aerobic fitness capability. It is one of the most challenging assessments that will not suit beginners.

5. Conclusions and Recommendation

This study shed light on the potential of the three field tests as assessment tools for measuring the cardiovascular endurance of the college freshmen students of the Physical Fitness and Movement Enhancement (HKD-01) course. Specifically, this identified (1) the validity of the three tests as assessment tools for gauging cardiovascular fitness and (2) their reliability as shown in the correlation between the directly measured and predicted VO₂ max. The results of the statistical analyses revealed that the fitness tests were found valid and reliable, thus, equally practical tools for assessing university non-athlete students' cardiovascular endurance. However, their level of difficulty should be considered. The step test is used for a starter-level as it is the least complicated of the three to execute. The jumping rope test is intermediate level as it takes a lot of practice to perform correctly but not as difficult as the beep test which is more recommended for advanced level and strictly considers the physical capacity and condition of the participants.

It is then recommended that PE professors gauge the cardiovascular endurance of each student using these tests

if they are implementing new courses or activities that require physical performance. It is also suggested that professors also test the validity and reliability of their assessment tools in an effort to ensure quality PE instruction. It is best to administer evidence-based tests with a keen consciousness of the right cardiovascular assessment procedure to ensure that the professors keep track of the students' fitness progress while keeping in mind the wellbeing of the students as well.

Future researchers may further examine the validity and reliability of other common assessment tools used in PE classes and consider other anthropometric variables as predictors of low or high physical performance of students in these three tests.

REFERENCES

- [1] Jess C., Melissa B., Christopher B., Z. Papalia, "College Students' Work Habits are Related to Physical Activity and Fitness," *International Journal of Exercise Science*, vol. 10, no. 7, p. 1009-1017, 2017.
- [2] Peter P., Carol B., Sonya M., J. Thayer, "Trends in Body Fat, Body Mass Index, and Physical Fitness among Male and Female College Students," *Nutrients* vol. 2, pp. 1075-1085, 2010.
- [3] Justin L., Mark T., Luc L. Tim O., G. Tomkinson, "International Variability in 20 M Shuttle Run Performance in Children and Youth: Who are the Fittest From a 50-Country Comparison: A Systematic Literature Review with Pooling of Aggregate Results," *British Journal of Sports Medicine*, vol. 52, no. 4, pp. 1-12, 2016.
- [4] Camila Angelica Asahi M., Bruna Camilo T., Rodrigo Dutra B., Rosana Lisboa M. Sandara Lia A., H. Monteiro, "Health-Related Physical Fitness among Undergraduate Students in Physical Education," *Journal of Physical Education*, vol. 29, no. 8, pp. 1-9, 2018.
- [5] Heildenberg D. G. Soriano, "Effectiveness of the Multistage Jumping Rope Program in Enhancing the Physical Fitness Levels among University Students," *International Journal of Human Movement and Sports Sciences*, vol. 8, no. 5, pp. 235 – 239, 2020. DOI: 10.13189/saj.2020.080511.
- [6] Thomas F. Paul I., W. Roth, "Fit & Well: Core Concepts and Labs in Physical Fitness and Wellness," (12th ed), McGraw-Hill, 2017.
- [7] Derya Ozer K. Irem D. Gul B., Selma K., F. Colakoglu, "The Effects of Rope or Weighted Rope Jump Training on Strength, Coordination, and Proprioception in Adolescent Female Volleyball Players," *The Journal of Sports Medicine and Physical Fitness*, vol. 51, pp. 211-9, 2011.
- [8] Sadi P., "Effects of 7 Weeks of Rope - Jump Training on Cardiovascular Endurance, Speed, Agility in Middle School Student Boys," *Sport Science*, vol. 6, no. 2, pp. 40 - 43, 2013.
- [9] Hamid A., Aliakbar J., H. Abdinejad, "A Comparison of Two Aerobic Training Methods (Running vs Rope Jumping) on Health-Related Physical Fitness in 10 to 12 Years Old Boys," *Physical Activity Review*, vol. 4, pp. 9-17, 2016.
- [10] Chao-Chien C., Y. Lin, "Jumping Rope Intervention on Health-Related Physical Fitness in Students with Intellectual Impairment," *The Journal of Human Resource and Adult Learning*, vol. 8, no. 1, pp. 56-62, 2012.
- [11] Richa Hirenda R. Priyanka S., M.P.S. Negi, "A Study on Cardiovascular Fitness of Sedentary College Students," *International Journal of Science and Research*, vol. 4, no. 6, pp. 109-112, 2015.
- [12] Emily K., Melanie S., R. Petrella, "Validation of the Step Test and Exercise Prescription Tool for Adults," *Canadian Journal of Diabetes*, vol. 38, pp. 164-171, 2014.
- [13] Jennifer C., Jonathan M., Yasmeen A., Jeremy J., Andrew L., Francesco C., Peter M. J. Thorn, "A Simple Step Test to Estimate Cardio-Respiratory Fitness Levels of Rheumatoid Arthritis Patients in a Clinical Setting," *International Journal of Rheumatology*, pp. 1-6, 2013.
- [14] Brooke S., Leona C. Naveen U., Sarah M., Amy S., Carolyn L., Danielle M., Mary Ann O., S. Narasimhan, "Submaximal Step-Testing as a Tool to Assess Cardiovascular Fitness in Youth Who are Obese," *Medicine and Science in Sports Exercise*, vol. 43, 2011.
- [15] Giorgos P., Elias Z., Dafni M., Athanasia S. Polyxeni A., C. Cooke, "Multi-Stage 20-M Shuttle Run Fitness Test, Maximal Oxygen Uptake, and Velocity at Maximal Oxygen Uptake," *Journal of Human Kinetics* vol. 41, pp. 81–87, 2014.
- [16] Daniel M. Pablo A., J. Viciano, "Criterion-Related Validity of the 20-M Shuttle Run Test for Estimating Cardiorespiratory Fitness: A Meta-Analysis," *Journal of Sports Science and Medicine*, vol. 14, no. 3, pp. 536-47, 2015.
- [17] Michael S., Stewart P., M. Bouffard, "Prediction of Maximal Aerobic Power from the 20-M Multistage Shuttle Run Test," *Canadian Journal of Applied Physiology*, vol. 28, pp. 272-82, 2003.
- [18] Michael H., Meegan F., Catherine L., Nick D., Jeremy S., N. Kimber, "Measurement of Cardiorespiratory Fitness in Children from Two Commonly Used Field Tests after Accounting for Body Fatness and Maturity," *Journal of Human Kinetics*, vol. 40, pp. 83-92, 2014.
- [19] Vladimir P., Arlene N., C. Boreham, "A Multistage Shuttle Run: A Predictor of Running Performance and Maximal Oxygen Uptake in Adults," *Brit. J. Sports Med*, vol. 21, no. 4, pp. 163-165, 1987.
- [20] Cristina C., Francisco A., Guillermo S., Jose M., Borja M., Manuel H., Manuel A., David J., Pedro F., Jonatan R., F. Ortega, "Reliability of an Adaptation of the 20m Shuttle Run Test to be Used in Preschool Children: The PREFI 20m Shuttle Run Test," *Rev Andal Med Departe*, vol. 8, no. 1, p. 23, 2015.
- [21] Luc L., Daniel M., Christine G., J. Lambert, "The Multistage 20 Metre Shuttle Run Test for Aerobic Fitness," *Journal of Sports Sciences*, vol. 6, pp. 93-101, 1988.
- [22] Amit B., "Validity of 20-Meter Multistage Shuttle Run Test for Estimation of Maximum Oxygen Uptake in Male

- University Students,” *Indian Journal of Physiology and Pharmacology*, vol. 55, no. 3, pp. 221-226, 2011.
- [23] Amit B., “Validity of 20-Meter Multistage Shuttle Run Test for Estimation of Maximum Oxygen Uptake in Female University Students,” *Indian Journal of Physiology and Pharmacology*, vol. 57, pp. 77-83, 2013.
- [24] Melanie J., E. Mascardo, *The DLSU FWTEAMS Basic Basketball Instructional Manual*, C & E Pub, 2008.
- [25] Slaughter M., Tim L., Richard B. Craig H. C., Richard S., Marta V.L., D. Bembem, “Skinfold Equations for Estimation of Body Fatness in Children and Youth,” *Human Biology*, vol. 60, no. 5, 709-723, 1988. <http://www.jstor.org/stable/41464064>
- [26] Christine V., G. Sandercock, “Does the Twenty Meter Shuttle-Run Test Elicit Maximal Effort in 11-To 16-Year-Olds?” *Pediatric Exercise Science*, vol. 21. pp. 55-62, 2009. 10.1123/pes.21.1.55.
- [27] Patrick T. “How to Use Difference Plots in Quantitative Method Comparison Studies,” *Annals of Clinical Biochemistry*, vol. 43, 124 – 129, 2006.
- [28] Meagan P., Jefferson, Sean J., John K, Kevin L. Greg M., C.B. Winnard. “Method Comparison: A Different Approach. *Annals of Clinical Biochemistry*,” vol. 29, pp. 556 – 560, 1992.
- [29] Bland-Altman plot. MEDCALC, <https://www.medcalc.org/manual/blandaltman.php> (accessed Dec. 23, 2020)
- [30] Catey B., Correlation, Agreement, and Bland-Altman Analysis: Statistical Analysis of Method Comparison Studies, 2008. doi:10.1016/j.ajo.2008.09.032
- [31] Davide G., “Understanding Bland-Altman Analysis,” Lesson in Biostatistics, 2015. https://www.biochemia-medica.com/assets/images/upload/xml_tif/Giavarina_Davide-Understanding_Bland_Altman_analyses.pdf