ISSN: 2349-7637 (Online)
RESEARCH HUB International Multidisciplinary Research Journal
https://rhimrj.co.in/ [Peer-Reviewed Journal]

Reliability of Field Based Health Related Physical Fitness Tests in Sri Lankan School Settings

[*]K.S.H.M.V.W.W.Senevirathne

Department of Education, Faculty of Arts, University of Peradeniya, Sri Lanka

ARTICLE DETAILS

Article History
Published Online: 17 April 2021

Keywords
Health Related Physical Fitness.

[*]Corresponding Author
Email: *walters@pdn.ac.lk*

ABSTRACT

Health-related fitness is demonstrated by being able to perform daily activities with vigor and by traits and capacities associated with a low risk for the development of chronic diseases and premature death. Physical fitness testing in the field simple tests to evaluate different components of fitness. Objective of this study was to determine reliability if the BMI test, waist circumference, one mile walk test,4x10m shuttle run test, sit and reach test, standing long jump test and push-ups test in Sri Lankan Schools settings. A quantitative approach was used in this study. Physical fitness assessments were implemented to measure health related physical fitness of 114 which was randomly selected sample in the Kandy District. In this study used three statistical methods to determine the reliability of tests, Inter-Trials agreement of the Bland –Altman Plot diagrams, One –way ANOVA for repeated measures and Cronbach's Alpha co-efficient values. Bland-Altman plots graphically shows the reliability patterns, in terms of systematic error (bias or mean inter trial differences) and random error (95 % limits of agreement), the heteroscedasticity analysis showed no positive association between inter-trial differences and inter-trial mean all the elements of the tests. Findings of this study relating to the reliability suggest that the health-related fitness tests, which included the one Mile walk, 4x10 m shuttle run, push- ups and standing long jump tests, sit and reach test and measures of weight, stature, and waist circumference administered by PE teachers in the Kandy district in Sri Lankan school setting can be considered as a reliable test.
According to the George and Mallery scale, alpha value for the whole sample of 0.709 indicates that internal consistency co-efficient of this study is "Good". However, as depicted above in Table 3 alpha values, for waist measurement and sit and reach test recorded less than 0.7. But it is also at the "acceptable" level. Therefore, reliability of field-based health related physical fitness tests in Sri Lankan School Settings is depicted at significantly acceptable level.

1. Introduction

Physical fitness is a set of attributes that people have, or achieve, related to their ability to perform physical activity (Bouchard C, &Sheppard RJ, 1994). Health-related physical fitness consists of those components of physical fitness that are associated with health. Health-related fitness is demonstrated by being able to perform daily activities with vigor and by traits and capacities associated with a low risk for the development of chronic diseases and premature death (US Department of Health and Human Services, 2009). Physical fitness testing in the field typically involves the administration of a battery of simple tests to evaluate different components of fitness (ACSM, 2000). The most common field tests of physical fitness for children include the following elements:Aerobic fitness/capacity tests: distance/timed walks/runs; step tests; multistage fitness test,Muscular strength/endurance: sit ups/curl ups and/or the progressive abdominal sit up (curl) test; pull ups and/or modified pull ups; push-ups,Flexibility: sit and reach; shoulder stretch; arm lift,Body Composition: body mass index (BMI); skinfold thicknesses; girth measures. (ACSM, 2000)

Reliability is one of the most important elements of test quality. It has to do with the consistency, or reproducibility, or an examinee's performance on the test. For example, if you were to administer a test with high reliability to an examinee on two occasions, you would be very likely to reach the same conclusions about the examinee's performance both times. A test with poor reliability, on the other hand, might result in very different scores for the examinee across the two test administrations. If a test yields inconsistent scores, it may be unethical to take any substantive actions on the basis of the test. There are several methods for computing test reliability including test-retest reliability, parallel forms reliability, decision consistency, internal consistency, and inter-rater reliability. For many criterion-referenced tests decision consistency is often an appropriate choice (Professional Testing Inc. 2006). Objective of this study was to determine reliability if the BMI test, waist circumference, one mile walk test,4x10m shuttle run test, sit and reach test, standing long jump test and push-ups test in Sri Lankan Schools settings.

2. Materials and Methods

A quantitative approach was used in this study. Physical fitness assessments were implemented to measure health related physical fitness of 114 which was randomly selected sample in the Kandy District. A simple random sampling method was adopted to select this student sample aged between 11 years and 17 years. To ensure standardization of data collection six physical education teachers thoroughly trained as data collectors through training workshops. All the calculations were performed using SPSS vs. 17 for windows. Several statistical methods have been traditionally used for assessing reliability. In this study researcher used three statistical methods to determine the reliability of tests.

1. Inter-Trials agreement of the Bland –Altman Plot diagrams
2. One –way ANOVA for repeated measures
3. Cronbach's Alpha co-efficient value

The Pearson's correlation coefficient (r) depends greatly on the range of values of the sample (high inter-individual variation increases the correlation) (Atwater et al, 1990). A high correlation between repeated scores reflects well the stability of position or rank order within a particular sample. However, it is not possible to assess the systematic error which is also known as bias. According to Atwater et al (1990) the Coefficient of Variation (CV) provides useful information about random error of the measure, but not about the systematic error (bias). The Intra-class Correlation Coefficient (ICC) is another measure of correlation that equals 1 if the two measurements are identical in every case. It is suggested that an ICC between 0.7 and 0.8 is considered 'questionable',> 0.9 is considered 'high', and close to 1indicates 'excellent' reliability (Vincent ,1994). ICC is an appropriate overall summary measure of agreement because it reflects both systematic bias and random error (Rothwell, 2009). ICC is however affected by the sample range, so that a high correlation may still indicate an acceptable measurement error. In addition, it does not give any information on variation in agreement with the size of the measurement. When the amount of variance between test and retest increases as the magnitude of the measurement increases (or decreases), the data are said to be heteroscedastic (Ruiz et al, 2011) (see details in Chapter on methodology), an issue that is rarely mentioned in sports and exercise reliability studies (Atkinson&Nevill, 1998).

According to the literature (Atkinson&Nevill, 1998), paired statistics or ANOVA for repeated measures also provides important information in reliability studies, mainly to find out to whether he results of test and retest are significantly different. Furthermore, the Bland-Altman method with limits of agreement has been proposed for a meaningful and useful interpretation of reliability, as it shows the measurement error schematically (both bias and random error) and helps to identify the presence of heteroscedasticity (Vincent, 1994, Atkinson&Nevill, 1998).

According to the above-mentioned statistical data analysis methods the researcher has selected the ANOVA for repeated measures and Bland-Altman plot diagrams for the analysis of reliability of the health-Related physical fitness tests. Furthermore, to strengthen the analysis of reliability of this test Cronbaach's alpha co-efficient method was used in this study. Test and retest (hereafter called T1 and T2) were compared between boys and girls by means of one -way repeated measures analysis of variance (ANOVA), with sex and age group as fixed factors. The analyses were performed for girls and boys together. However, age group-specific effect on reliability was found in the study tests (all P < 0.05). Potential systematic error between test and retest (also called bias) was analyzed by means of repeated measures ANOVA. The differences between T1 and T2 scores was also examined through different error measures. Suppose that *N* cases are available to evaluate the error measurements, where ŷ is T2 and y is T1 and mentioned in Chapter three. The sum of squared errors (SSE) was also calculated as mentioned in the chapter three of this thesis.

Moreover, the inter-trials agreement was also examined graphically by plotting the difference between each pair of measurements against their mean, according to the Bland and Altman approach. The 95 % limits of agreement for all the physical fitness variables were calculated as the inter trials mean difference± 1.96 SD of the inter trial differences. As mentioned in the methodology chapter test and retest (T1 and T2) was held in a 6 days' time gap. The presence of heteroscedasticity was examined by using "One Way ANOVA" in order to determine whether the absolute inter-trials difference was associated with the magnitude of the measurement (i. e., inter-trials mean). For that purpose, age group-specific quartiles of the inter-trials mean were established.

A significant association (P < 0.05) would confirm heteroscedasticity, which means that the inter-trials difference of a test would differ with the physical fitness level of a group. All analyses were performed using SPSS v.17 software for Windows. For all analyses, the significant level was set at 5 %.

3. Results

3.1The inter-trials agreement of the Bland-Altman plots

The Bland –Altman plot of the tests in children, the central line represents the mean differences between the second trial(T2) and the first trail (T1); the upper and lower lines represents the upper and lower 95% limits of the agreement (means difference ±1.96 SD of the differences) respectively.

Figures 1 to 8 illustrate the Bland –Altman plots for the reliability patterns, in terms of systematic error (bias or mean inter trial differences) and random error (95 % limits of agreement) of all the elements of the tests are as follows.

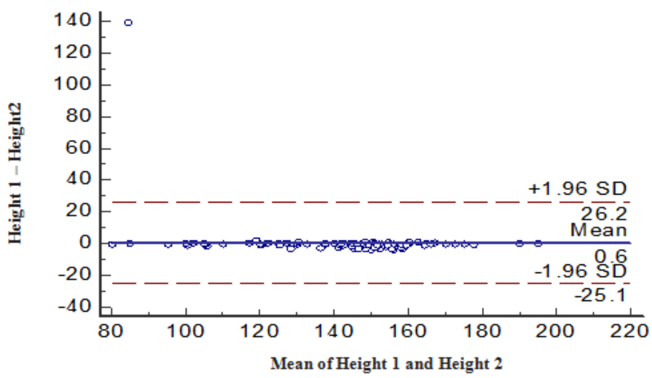

Figure 1: **Bland - Altman plots of Height measurement**

The above Figure 1 shows mean of the height of the sample between 80 cm and 190 cm. The inter-trial differences fall between -10 to and +10.The heteroscedasticity analysis showed no positive association between inter-trial differences and inter-trial mean of height measurements showed in the above Bland-Altman plot diagram.

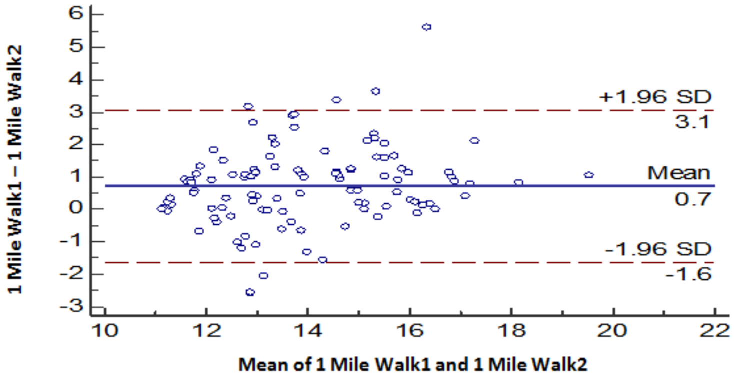

Figure2: **Bland - Altman plots of 1 Mile walk measurement**

The above Figure 2 shows means of the one mile walk test of the sample between 10 minutes and 18 minutes. The inter-trial differences fall between -2.5 to and +5.5. Although some values deviate from the acceptable limits but most of the inter trail values fall in the acceptable limits. The heteroscedasticity analysis showed no positive association between inter-trial differences and inter-trial mean of one mile walk test showed in the above Bland-Altman plot diagram.

Figure 3 shows means of the push-up test of the sample between 1push up and 38 push -ups. The inter-trial differences fall between -25 to and +25.Although very little values are deviate from the acceptable limits, most of the inter trail values falls in the acceptable limits. The heteroscedasticity analysis showed no positive association between inter-trial differences and inter-trial mean of push-up test showed in the above Bland-Altman plot diagram.

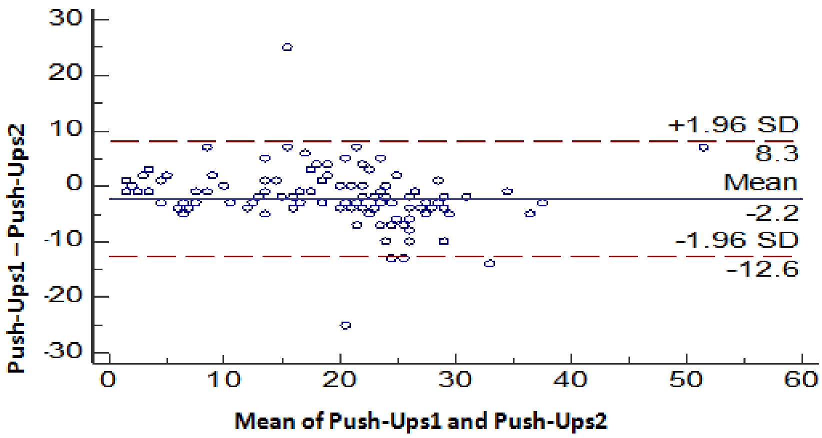

Figure 3: **Bland - Altman plots of push-ups measurement**

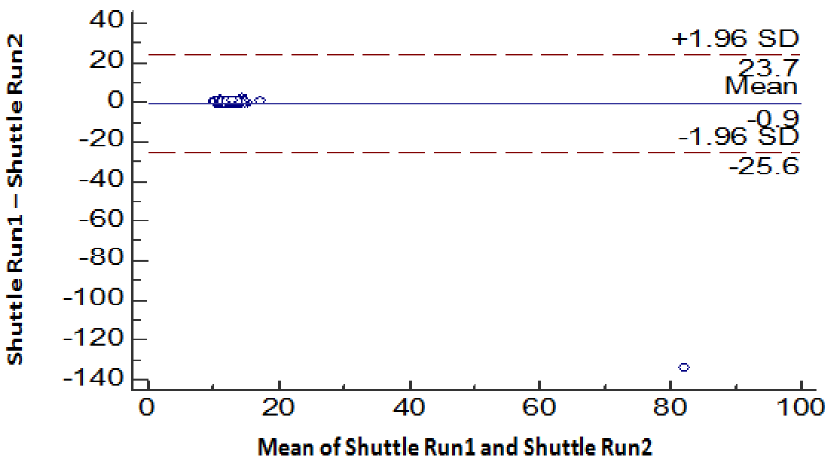

Figure 4: **Bland - Altman plots of 4x10 m shuttle run measurement**

The above Figure 4 shows mean of the 4x10m shuttle run test of the sample between 10 seconds and 18 seconds. The inter-trial differences falls between -5 to and +5.This indicates very small range .This plot diagram shows that all the inter trail values fall significantly within the acceptable limits. The heteroscedasticity analysis showed no positive association between inter-trial differences and inter-trial mean of 4x10m shuttle run test showed in the above Bland-Altman plot diagram.

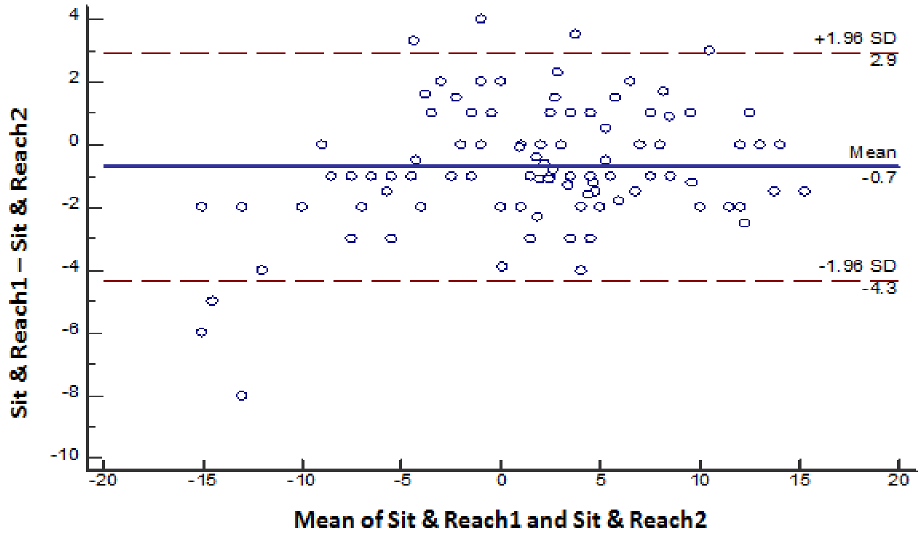

Figure 5: **Bland - Altman plots of sit & reach measurement**

The above Figure 5 shows mean of the sit and reach test of the sample between -15 cm and +15 cm. The inter-trial differences fall between -8 to and +4.Although very little values are deviating from the acceptable limits, but most of the inter trail values fall within the acceptable limits. The heteroscedasticity analysis showed no positive association between inter-trial differences and inter-trial mean of sit and reach test showed in the above Bland-Altman plot diagram.

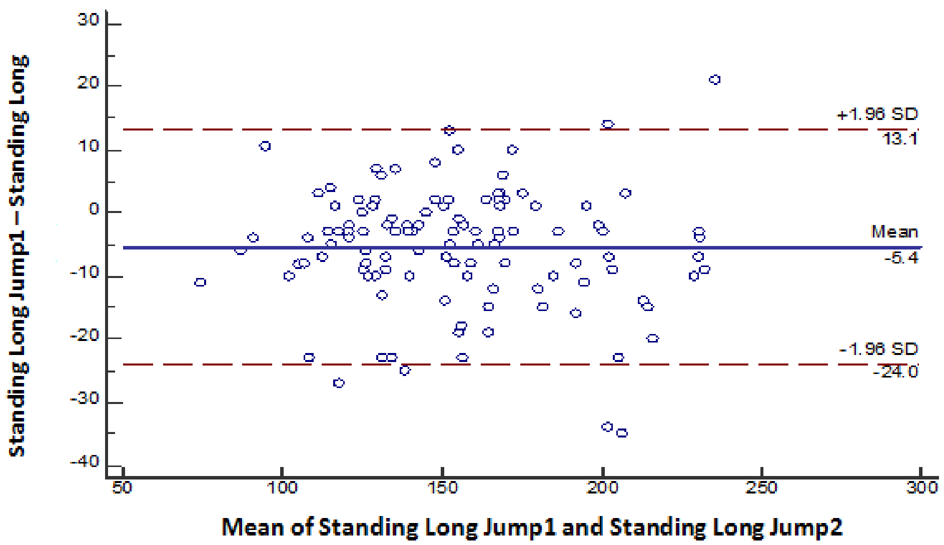

Figure 6: **Bland - Altman plots of standing long jump measurement**

Figure 6 above shows the mean of the standing long jump test of the sample between 75 cm and 230 cm. The inter-trial differences fall between -35 to and +20.Although very small numbers of values deviate from the acceptable limits and scattered, but most of the inter trail values fall within the acceptable limits. The heteroscedasticity analysis showed no positive association between inter-trial differences and inter-trial mean of standing long jump test showed in the above Bland-Altman plot diagram.

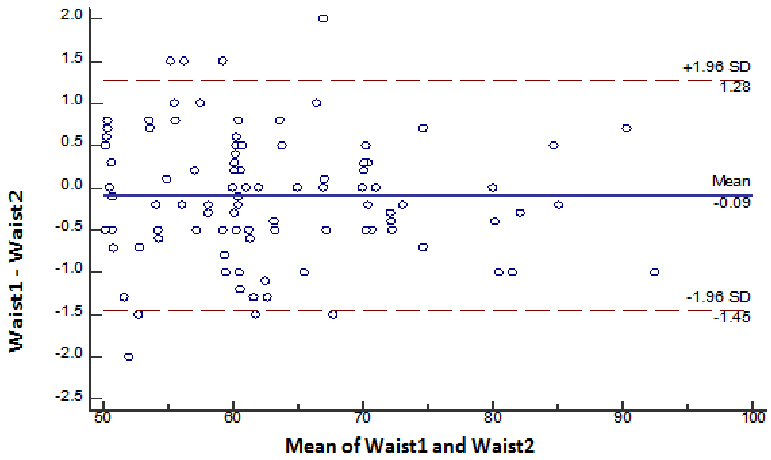

Figure 7: **Bland - Altman plots of waist measurement**

The above Figure 7 shows the mean of the waist measurement test of the sample between 50 cm and 95 cm. The inter-trial differences fall between -2 to and +2. Although very small numbers of values deviating from the acceptable limits and scattered, but most of the inter trail values fall within the acceptable limits. The heteroscedasticity analysis showed no positive association between inter-trial differences and inter-trial mean of waist measurements test showed in the above Bland-Altman plot diagram.

The bellow Figure 8 shows mean of the weight measurement of the sample between 25 Kg seconds and 75Kg. The inter-trial differences fall between -5 to and +5. This indicates very small range. This plot diagram shows that all the inter trail values fall significantly within the acceptable limits for all but one plot. The heteroscedasticity analysis showed no positive association between inter-trial differences and inter-trial mean of weight measurement showed in the above Bland-Altman plot diagram.

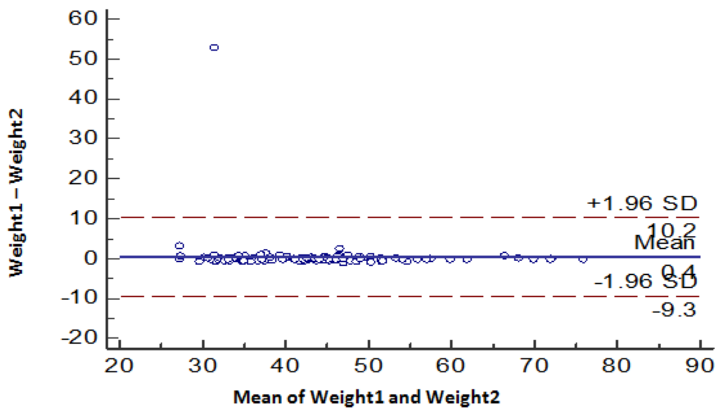

Figure 8: **Bland - Altman plots of weight measurement**

According to the above-mentioned Bland-Altman plots Figures 1 to 8 graphically shows the reliability patterns, in terms of systematic error (bias or mean inter trial differences) and random error (95 % limits of agreement), in those measurements which presented a significant inter-trial difference. The heteroscedasticity analysis showed no positive association between inter-trial differences and inter-trial mean showed in the above Bland-Altman plots in all the elements of the tests. It was revealed that all these plot diagrams depict that the tests which was implemented in this study was reliable to use secondary schools in the Kandy district in Sri Lanka.

3.2 One –way ANOVA for repeated measures

Mean and SD values for the test and retest trials according to age groups are shown in Table 1.

Table 1: **Mean and SD values for the T1 and T2 for different fitness components**

Test	Whole sample test 1 Mean and SD	Whole sample test 2 Mean and SD	Test1 Boys Mean	Test2 Boys Mean	Test 1 Girls Mean	Test2 Girls Mean
Weight	43.24 ± 9.79	42,81 ± 10.49	43.7983	42.9627	42.6600	42.6273
Height	142.4 ± 20.028	141.9 ± 24.26	135.2898	135.4907	150.2000	148.8073
Waist	63.32 ± 8.98	63.41 ± 9.01	64.3186	64.2985	62.2636	62.4673
Push-Ups	18.56 ± 8.86	20.74 ± 10.14	15.1695	15.6102	22.2000	26.2545
Standing Long Jump	152.01 ± 35.53	157.43 ± 36.28	172.4746	179.0339	130.0709	134.2636

Sit & Reach	1.34 ± 6.8	2.0465 ± 6.34	0.9085	1.8085	1.8200	2.3018
4x10m Shuttle Run	12.6 ± 1.23	13.54 ± 12.86	11.9415	11.6008	13.3233	15.6347
1 Mile Walk	14.28 ± 1.95	13.56 ± 1.69	13.0697	12.6690	15.5976	14.5327

To test the mean differences of the values of the test and retest were compared between boys and girls by means of one- way repeated measures analysis of variance (ANOVA) with sex and age group as fixed factors. The analyses were performed for girls (n=54) and boys (n=60) together (total sample, n=114).

Table 2: **Analysis of Variance of the test and retest**

		Sum of Squares	**df**	**Mean Square**	**F**	**Sig. P value**
weight	Between Groups	11.453	1	11.453	0.112	0.738
	Within Groups	23061.615	226	102.043		Not Significant
	Total	23073.068	227			
Height	Between Groups	18.388	1	18.388	0.036	0.85
	Within Groups	116451.034	226	515.27		Not Significant
	Total	116469.422	227			
Waist	Between Groups	1672.938	1	1672.938	1.297	0.256
	Within Groups	291437.935	226	1289.548		Not Significant
	Total	293110.872	227			
Standing Long Jump	Between Groups	0.439	1	0.439	0.005	0.941
	Within Groups	18324.666	226	81.083		Not Significant
	Total	18325.106	227			
Push-Ups	Between Groups	271.934	1	271.934	2.997	0.085
	Within Groups	20503.693	226	90.724		Not Significant
	Total	20775.627	227			
4x10m Shuttle run	Between Groups	50.243	1	50.243	0.602	0.439
	Within Groups	18868.806	226	83.49		Not Significant
	Total	18919.049	227			
One Mile walk	Between Groups	11.642	1	11.642	3.847	0.053
	Within Groups	757.25	226	3.351		Not Significant
	Total	786.892	227			
Sit & Reach	Between Groups	27.79	1	27.79	0.634	0.427
	Within Groups	9904.268	226	43.824		Not Significant
	Total	9932.058	227			

According to table 2, the main findings of this study relating to the reliability suggest that the health-related fitness tests, which included the one Mile walk, 4x10 m shuttle run, push- ups and standing long jump tests, sit and reach test and measures of weight, stature, and waist circumference administered by PE teachers in the Kandy district in Sri Lankan school setting can be considered as a reliable test. According to the above output of the ANOVA table (Table 2) it was revealed that the studied fitness tests also have an acceptable level of reliability when administered by a trained staff in Sri Lankan secondary schools. No significant differences were reported between test and retest in push -ups (upper body strength) in Sri Lankan secondary school children in the Kandy district. It is also revealed that the standing long jump (lower body strength) tests are reliable in Sri Lankan secondary school children. No significant differences between test and retest were reported in the one Mile walk/ run test in Sri Lankan secondary school children aged 11 – 17 years. However, it was observed that the lowest reliability recorded in the one mile walk test among these test elements in the amended extended ALPHA health-related fitness test battery. Regarding anthropometric assessment, the error has been traditionally considered unavoidable and has been frequently studied by means of technical error of measurement (Ruiz and colleagues, 2011). The inter- and intra-observer measurement errors for waist circumferences can be minimized by paying close attention to every aspect of the data collection process. With the aim of making the data as close to the reality as possible, researcher trained PE teachers to administer the measurement very accurately in the school setting, during the training workshops for PE teachers. It was observed that the reliability was high in all the tests (range 0.053 - 0.941) except in one mile walk test and push –ups test in Sri Lankan secondary school children (P values were 0.053 and 0.085 respectively). Possible reasons for that are discussed in discussion part.

3.3 Cronbach's Alpha value of the reliability of tests in Sri Lankan School Settings.

There are several methods to examine the internal consistency co-efficient. Cronbach’s alpha co-efficient method was used to examine reliability of this study. Cronbach’s alpha reliability co-efficient normally ranges between 0 and 1. However, there is actually no lower limit to the co-efficient. The closer the Cronbach’s alpha is to 1.0 the grater the internal consistency of the items in the scale. Based upon formula α= rk/ [1+ (k-1) r] where k is the number of items considered and r is the mean of the scale. The mean inter-item correlations the size of alpha is determined by both the number of items in the scale and the mean inter- item correlations. George and Mallery (2003) provide the following rules of thumb for comparing the values of alpha (α): “>0.9- Excellent, >0.8 – Good, >0.7- Acceptable, > 0.6 – questionable, >0.5 –poor and <0.5 unacceptable”. Cronbach’salpha co-efficient was determined in this study using T_1 and T_2 test scores for the whole sample and for each test elements.

Table 3: **Cronbach's Alpha values of the reliability test**

Item	**N of Items**	**Cronbach's Alpha**
For the whole sample	*16*	*0.709*
Weight	2	0.711
Waist	2	0.695
Height	2	0.702
Standing long jump	2	0.713
Push-up	2	0.700
Sit & Reach test	2	0.699
4x10m shuttle run	2	0.708
One mile walk	2	0.702

According to the George and Mallery (2003) scale, alpha value for the whole sample of 0.709 indicates that internal consistency co-efficient of this study is “Good”. However, as depicted above in Table 3 alpha values, for waist measurement and sit and reach test recorded less than 0.7. But it is also at the “acceptable” level. Therefore, reliability of field-based health related physical fitness tests in Sri Lankan School Settings is depicted at significantly acceptable level.

4. Discussion

These findings differ with similar studies especially in European Union studies. For example, the original ALPHA – FIT battery used 20 m Shuttle run test to measure cardiorespiratory fitness, instead of the one mile walk test that is used in this study (Ruiz and colleagues, 2011). They recorded higher reliability values than this study in 20 m Shuttle run test (cardiorespiratory fitness). Furthermore, they reported an intra-class correlation of 0.91 for repeated measures on 21 boys and girls 14-18 years of age. Furthermore, Table 4 shows R values of some selected studies.

Table 4: **Comparison of reliability values with other studies**

Researchers	**Male (n)**	**Female (n)**	**Age range**	**R**
Cureton (1995)	490	263	8-25 yrs	R=0.72
Rowland et al. (1999)	36	0	6^{th} grade	R =0.77
Current study	60	54	11-18 yrs	R=0.427-0.91

Current study reported that weight measurements, height measurements, standing long jump measurements, sit & reach test and 4x10 m shuttle run test measurements had a good reliability values for these test elements. They were 0.738, 0.85, 0.941, 0.427 and 0.49 respectively. This study showed a relatively low reliability in the one mile walk test and push -ups. Therefore, to obtain good reliability values for One Mile walk test and push- ups measurements, more attention is required especially for the younger children because of the importance of proper positioning and technique of 90-degree push-up and in order to obtain an accurate measurement and highly motivated commitment in walking. Several studies (Ruiz and colleagues, 2011) have indicated that the reliability of push- up measurement may improve in children when two or more measures are obtained with certain time gap and averaged as done in this study. However, reliability of the push up test recorded comparatively low value in this study. Meanwhile, comparatively low reliability in the one mile walk test could be due to the fact that this measurement was performed by persons with different levels of commitment and under different weather conditions. It is also the pacing may be a problem with younger children (starting too fast and getting exhausted early).

With the one-mile run, the prediction of aerobic capacity depends on other factors (including the child's BMI). Therefore, it is not possible to provide a direct goal time for which students should aim. Students should be encouraged to cover the distance as quickly as possible. Similarly, with the One-Mile Walk test it is not possible to produce estimated goal times. This is because performance on the assessment depends on the child's heart rate relative to the time it took to complete the walk. One student may prefer a faster walking pace while others may use a slower pace. An advantage of the walk/run test is that it is possible to estimate aerobic capacity regardless of the pace that is chosen to complete the walk. There was moderate evidence about the level of reliability of the 1-milerun / walk test due to the low number of studies (Artero et al, 2011). Therefore, more days of training for students could be needed so that the One mile walk test would be easier for them to perform and accordingly more reliable.

Although there were no significant differences between test and retest measurements in stature and, waist circumference for Sri Lankan secondary school children, the mean inter-trial difference for these measurements was nearly 0.5 cm. This suggests that despite finding no significant differences between test and retest, from a practical point of view these tests can be used in the school setting to evaluate children and adolescents. Cronbach's alpha co-efficient values were used to strengthen the statistical evidence and to justify the reliability of this study. It is accepted statistical evidence which is used in the quantitative data analysis. Therefore, the researcher has decided to use this method of analysis in addition to ANOVA.

5. Conclusions

From a practical point of view, this study proposes a health-related fitness test battery that is reliableto be administered by PE teachers in the school setting. Given the well-documented evidence about physical fitness as a marker of health status in children and adolescents the provided test will help to identify children and adolescents with low physical fitness levels, which will be useful for health promotion and disease prevention programs. In conclusion, the present study indicates that health-related fitness tests including the 4x10 m shuttle run, Push- ups test, standing long jump tests, sit and reach test, 1 mile walk test and measures on weight, stature, waist circumference, administered by PE teachers is reliable to be performed in the Sri Lankan school setting in the Kandy district.

References

1. American College of Sports Medicine (ACSM), (2000). *Exercise testing and prescription for children, the elderly, and pregnant women.* In: ACSM's Guidelines for Exercise Testing and Prescription. Sixth edition, Lippincott Williams &Wikins, pp. 217-234.
2. Artero, E.G., Espana-Romero, V., Castro-Pinero, J., Ortega, F.B., Suni, J., Castillo-Garzon M.J., Ruiz, J.R. (2011). Reliability of field-based fitness tests in youth. Int J Sports *Med*; 32, pp.159-169.
3. Atkinson, G., Nevill, A.M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to Sports Medicine. *Sports Med,* 26, pp.217-238.
4. Atwater, J. A., Wisdom, R., and Verma, I.M. (1990). "Regulated mRNA stability." *Annu Rev Genet* 24, pp.519-41.
5. George, D., and Mallery, P. (2003). SPSS. *For windows step by step: A simple guide and reference 11.0*, Boston: Allyn and Bacon
6. Professional Testing. (2006). Building high quality examination programs: The Validity, Professional Testing, Inc.
7. Ruiz, J.R., Castro-Pinero, J., Espana-Romero, V., Artero, E.G., Ortega, F.B., Cuenca, M.M., Jimenez-Pavon, D., Chillon, P., Girela-Rejon, M.J., Mora, J., Gutierrez, A., Suni, J., Sjostrom, M., Castillo, M.J.(2011).Field based fitness assessment in young people: The alpha health related fitness test battery for children and adolescents. *Br J Sports Med*; 45, pp. 518-524.