Equivalence Reliability and Convergent Validity of Percent Body Fat Prediction Equations

Peter D. Hart[1,2,*]

[1]Health Promotion Program, Montana State University - Northern, Havre, MT 59501
[2]Kinesmetrics Lab, Montana State University - Northern, Havre, MT 59501
*Corresponding author: peter.hart@msun.edu

Abstract **Background**: The fitness professional may often benefit from the use of a simple equation in determining a health outcome for an individual in lieu of a more complicated or expensive procedure. Therefore, the purpose of this study was to examine the reliability and validity of several standard prediction equations for percent body fat (PBF). **Methods**: Data used for this study came from a body composition assessment of N = 131 college students. Five different PBF prediction equations were used, with body mass index (BMI), age, and sex as inputs for each (PBFEQ1 thru PBFEQ5). Additionally, PBF using a bioelectric impedance (BIA) handheld device (PBFHH) was measured for each participant. Equivalence reliability was examined across the five PBF prediction equations using different analysis of variance (ANOVA) models of the intraclass correlation coefficient (ICC). Convergent validity between the prediction equations and PBFHH was determined by examining Pearson correlation coefficients and Bland and Altman limits of agreement (LOA). Reliability and validity was also examined for obesity classification using the Kappa statistic. **Results**: Reliability across the five PBF prediction equations was excellent for all ICC models in both female (ICCs > .985) and male (ICCs > .976) analyses. PBFHH scores adequately converged with scores from each prediction equation in both female (rs > .913) and male (rs > .817) analyses. LOA between PBFHH and PBFEQ5 indicate small to moderate bias of 4.0 ± 5.1% and 4.7 ± 7.9% in female and male analyses, respectively. Finally, reliability and validity of the prediction equations to classify participants into obese and non-obese categories ranged from moderate to almost perfect. **Conclusion**: This study provides psychometric evidence supporting the use of PBF prediction equations in a college student population.

Keywords: Body composition (BC), Percent body fat (PBF), obesity, prediction equations

Cite This Article: Peter D. Hart, "Equivalence Reliability and Convergent Validity of Percent Body Fat Prediction Equations." *American Journal of Sports Science and Medicine*, vol. 7, no. 2 (2019): 45-50. doi: 10.12691/ajssm-7-2-4.

1. Introduction

The measurement of an individual's percent body fat (PBF), as often performed in applied exercise science, can be achieved using both laboratory and field-based techniques. Laboratory methods, such as hydrostatic weighing, dual-energy x-ray absorptiometry (DEXA), and air displacement plethysmography (BOD POD), are reliable techniques but are both time consuming and expensive [1]. Field-based methods, such as bioelectrical impedance (BIA) and skin-fold thickness assessments, are more practical but include greater amounts of measurement error due to factors such as subject characteristics and technician training [2]. The use of PBF prediction equations that utilize easy to obtain inputs (predictor variables) is another option for both the exercise science researcher and fitness practitioner. Several PBF prediction equations, utilizing body mass index (BMI) as the main input, have already been validated in general populations

[3,4]. However, little is known about the psychometric properties of such PBF prediction equations in generally healthy younger adults such as college students. Therefore, the purpose of this study was to examine the equivalence reliability and convergent validity of five previously validated BMI-based PBF prediction equations in a younger healthy adult population.

2. Methods

2.1. Participants and Design

Data for this research came from a campus-based fitness assessment where N=131 college students attending a rural public university volunteered to participate in a series of body composition (BC) tests. Students were recruited by campus flyers and word-of-mouth and were offered free evaluation in exchange for their participation. Study protocols were reviewed and approved by the university system's institutional review board (IRB).

2.2. PBF Assessment

PBF was estimated for each participant using five different previously validated prediction equations [5-9] (see Table 1). The same inputs were required for each PBF prediction equation, including BMI, age, and sex. BMI (kg/m^2) was assessed by measuring height with a wall-mounted stadiometer and measuring weight with an electronic floor scale [10]. PBF was assessed with an Omron handheld bioelectric impedance device and followed manufacturer's procedures [11]. Obese classification was determined using cutoff values of $\geq 25\%$ for male and $\geq 32\%$ for female, indicating obesity.

Table 1. PBF prediction equations used in the equivalence reliability and convergent validity study

Equation	Study	Intercept	BMI	Sex	Age
PBFEQ1	Deurenberg [5]	-5.4	1.20	-10.8	0.23
PBFEQ2	Deurenberg [6]	-8.0	1.29	-11.4	0.20
PBFEQ3	Gallagher [7]	-10.0	1.46	-11.6	0.12
PBFEQ4	Jackson [8]	-13.9	1.61	-12.1	0.13
PBFEQ5	Jackson [9]	-9.0	1.39	-10.3	0.16

Note. Age input as years. BMI input as kg/m^2. Sex input as 1=male and 0=female.

Table 2. Intraclass correlation coefficients (ICCs) for single score one-way and two-way ANOVA models used in the equivalence reliability study

ICC Type	Model	Model Theory	Measurement Routine	Reliability Definition$	Formula	Comments
1,1	One-way random effects	The measured values variable is the DV and the subjects variable is the only IV Trial and error variances are combined to form within subject variance Used when subjects and trials are not crossed (each subject has a different set of randomly selected trials)	Only 1 measurement taken in practice	Absolute agreement* ICC(1,1)a	$$\frac{MS_S - MS_W}{MS_S + (k-1)MS_W}$$	Use if we consider each subject was assessed by a different set of trials This ICC will be affected by large trial differences
2,1	Two-way random effects	The measured values variable is the DV and the subjects and trials variables are the IVs Trial and error variance terms are separate Includes only trial variance and not random error (because trials considered randomly selected)	Only 1 measurement taken in practice	Absolute agreement* ICC(2,1)a	$$\frac{MS_S - MS_E}{MS_S + (k-1)MS_E + \frac{k}{n}(MS_T - MS_E)}$$	Use if we care about generalizing to all similar trials This ICC will be affected by large trial differences
				Consistency (correlation)# ICC(2,1)c	$$\frac{MS_S - MS_E}{MS_S + (k-1)MS_E}$$	
3,1	Two-way mixed effects	The measured values variable is the DV and the subjects and trials variables are the IVs Trial and error variance terms are separate Includes only random error and not trial variance (because trials considered fixed effects and not randomly selected)	Only 1 measurement taken in practice	Absolute agreement# ICC(3,1)a	$$\frac{MS_S - MS_E}{MS_S + (k-1)MS_E + \frac{k}{n}(MS_T - MS_E)}$$	Use if we only care about these particular trials This ICC will not be affected by large trial differences
				Consistency (correlation)* ICC(3,1)c	$$\frac{MS_S - MS_E}{MS_S + (k-1)MS_E}$$	

Note. In this table, Trials represent the PBF prediction equations. IV is independent variable. DV is dependent variable. $Definition notation was operationalized for this study. Trial variance (MS_T) is considered systematic. When systematic variance is small, the mean square within (MS_W) from the one-way model and the mean square error (MS_E) from the two-way models (reflecting random error) are similar. The larger the subjects variance (MS_S) the higher the ICC. Consistency reliability considers if one set of scores (y) can equate to another set (x) plus a systematic error component (i.e., y = x + c) and is often referred to as norm-referenced reliability. Absolute reliability considers if one set of scores (y) equals another set (x) exactly (i.e., y = x) and is often referred to as criterion-referenced reliability. Absolute agreement ICCs include trial variance in the denominator. *Indicates Shrout and Fleiss cited models. #indicates McGraw and Wong cited models.

Table 3. Descriptive statistics on study variables and participant characteristics by sex

Sex	Variable	*Min*	*Mean*	*Median*	*SD*	*CV*	*Max*
Female (N=44)	Age	18.0	23.1	21.0	7.2	31.3	54.0
	BMI	18.8	24.4	23.9	3.5	14.3	33.2
	WC	60.0	75.0	72.7	8.2	10.9	94.0
	PBFHH	14.7	24.6	22.8	6.4	26.2	39.0
	PBFEQ1	21.3	29.2	27.8	5.0	17.2	43.9
	PBFEQ2	19.9	28.1	26.7	5.2	18.5	42.4
	PBFEQ3	20.0	28.8	27.6	5.5	19.1	42.4
	PBFEQ4	18.7	28.4	27.1	6.0	21.1	43.2
	PBFEQ5	20.0	28.6	27.3	5.4	18.7	42.3
Male (N=87)	Age	18.0	21.2	20.0	3.4	16.0	45.0
	BMI	20.8	27.7	27.0	4.6	16.8	41.6
	WC	71.5	86.6	84.0	9.6	11.0	122.0
	PBFHH	4.6	17.9	18.0	6.9	38.4	35.4
	PBFEQ1	13.4	21.9	21.1	5.8	26.4	38.3
	PBFEQ2	11.5	20.6	19.5	6.2	30.0	38.3
	PBFEQ3	11.7	21.8	20.5	6.9	31.6	41.9
	PBFEQ4	10.2	21.4	19.9	7.6	35.5	43.6
	PBFEQ5	12.9	22.6	21.3	6.6	29.2	41.7

Note. BMI is body mass index (kg/m^2). WC is waist circumference (cm). PBFHH is percent body fat (PBF) by handheld bioelectric impedance (%). PBFEQ1 thru PBFEQ5 are PBF values by prediction equations (%). CV is coefficient of variation. Variances were not significantly different across PBFEQ1-PBFEQ5 for either female (p=.781) or male (p=.150) data.

2.3. Statistical Analyses

Descriptive statistics were computed on all study variables including the five PBF estimates from the prediction equations. Three different analysis of variance (ANOVA) models were run using PROC GLM to allow for the proper mean square estimates needed in the PBF prediction equation reliability study. The first model was a one-way ANOVA used only to test the equality of variance assumption across PBF prediction equation groups (PBFEQ1-PBFEQ5). The second model was also a one-way ANOVA where a subjects variable was used as a class-level independent variable. The third model was a two-way ANOVA where both subjects and trials (PBF prediction equations) were used as class-level independent variables. Although all mean square estimates were available via the two-way ANOVA model, the statistical F tests needed for the one-way ANOVAs required a one-way ANOVA analysis. Equivalence reliability of the five PBF prediction equations was examined using the mean squares from the ANOVA analyses to compute several intraclass correlation coefficients (ICCs). There are ten commonly used forms of ICCs, five of which are appropriate when the assessment procedure requires a single measurement and five are appropriate when the assessment requires repeated measurements and the mean of those repeated measurements is the final value [12,13]. It was decided that only a single prediction equation would likely be used with any given application, and so, only single measure ICCs were computed in this reliability study. Each of the single measure ICC models with associated theory and formulae are described in brief (see Table 2). Additionally, using the mean squares from the ANOVA output, 95% confidence intervals (CIs) were constructed around each ICC [13]. Convergent validity for the PBF prediction equations and PBF from BIA was examined using both Pearson correlation coefficients and Bland and Altman limits of agreement (LOA) plots [14].

Finally, both equivalence reliability of the PBF prediction equations and their convergence with BIA were examined in terms of obesity classification using the Fleiss (reliability) and Cohen (validity) Kappa (κ) [15,16]. For ICC strength criteria, values < .50 indicate poor agreement, values between .50 and .75 indicate moderate agreement, values between .75 and .90 indicate good agreement, and values > .90 indicate excellent agreement [17]. For Kappa strength criteria, values ≤ 0 indicate no agreement, values between .01 and .20 indicate none to slight agreement, values between .21 and .40 indicate fair agreement, values between .41 and .60 indicate moderate agreement, values between .61 and .80 indicate substantial agreement, and values ≥ .81 indicate almost perfect agreement [18]. For Pearson correlation coefficient strength criteria, values > .70 indicate acceptable convergent validity evidence [19]. All analyses were reported separately for each sex. All analyses were performed using SAS version 9.4 [20].

3. Results

All N = 131 participants, N = 44 female and N = 87 male, had complete body composition data. Table 3 displays descriptive statistics for the study-related variables. As required for ANOVA-based analyses, PBF variances across the prediction equations were not significantly different for either female ($p = .781$) or male ($p = .150$) data. Table 4 displays the ANOVA tables for PBF measurements from the five different prediction equations by sex. Variance sources are for the two-way models, however, within-group variances are also shown for the one-way models. Although variance due to trials (systematic variance) was significant for both female ($MS_T = 146.1$, $p < .001$) and male ($MS_T = 217.5$, $p < .001$) models, variance between subjects was significant and much larger for female ($MS_S = 146.1$, $p < .001$) and male ($MS_S = 217.5$, $p < .001$) models.

Table 5 displays equivalence reliability statistics for the five PBF prediction equations by sex. As previously mentioned, only single measure ICCs were reported for both consistency and absolute measurement generalizations. ICCs were computed from the mean squares displayed in Table 4. For example, ICC(2,1)a used mean squares for subjects (MS_S), means squares for trials (MS_T) and mean squares for error (MS_E) in its computation. Specifically, ICC(2,1)a for females was computed with number of trials equal to $k = 5$ a sample size equal to $n = 44$ as follows:

$$ICC(2,1)a = \frac{MS_S - MS_E}{MS_S + (k-1)MS_E + \frac{k}{n}(MS_T - MS_E)}$$

$$ICC(2,1)a = \frac{146.08 - 0.28}{146.08 + (5-1)0.28 + \frac{5}{44}(7.85 - 0.28)} = .985.$$

Note that ICC(2,1)a is defined as an absolute reliability coefficient and therefore includes trial variance in the denominator, which in the above example, increases the size of the denominator and in turn decreases the reliability. The larger the trial variance becomes in any measurement scenario, the smaller the absolute reliability coefficient becomes and vice versa. The consistency reliability coefficients are not concerned with trial variance because mean differences across trials are less of a concern and relative position of subjects is more the concern.

Continuing with Table 5, although both one-way and two-way model ICCs are reported, it is likely only the two-way models are applicable. Regardless of reliability definition, all ICC models indicate excellent agreement in both female (ICCs > .985) and male (ICCs > .976) analyses. These results are also justifiable considering the lower limit (LL) of each ICC 95% CI in both female (LLs > .976) and male (LLs > .943) analyses. Table 6 displays convergent validity evidence for PBF measurements from the five prediction equations and PBF from handheld bioelectric impedance (PBFHH) by sex. Each prediction equation (PBFEQ1-PBFEQ5) successfully converged with PBFHH in both female (rs > .913, p < .0001) and male (rs > .817, p < .0001) analyses. Additionally, limits of agreement (LOA) plots between PBFHH and PBFEQ5 indicate small to moderate bias of 4.0 ± 5.1% and 4.7 ± 7.9% in female and male analyses, respectively (see Figure 1 and Figure 2). Specifically, PBF prediction equations resulted in greater PBF estimates as compared to BIA.

Table 4. Analysis of variance (ANOVA) tables for percent body fat (PBF) measurements from five different prediction equations by sex

Sex	Source	*df*	SS	MS	F	*p*
Female	Subjects	43	6281.29	146.08	525.38	<.0001
	Trials	4	31.39	7.85	28.22	<.0001
	Error	172	47.82	0.28		
	Total	219	6360.50			
	Within	176	79.21	0.45		
Male	Subjects	86	18708.26	217.54	419.92	<.0001
	Trials	4	188.25	47.06	90.84	<.0001
	Error	344	178.21	0.52		
	Total	434	19074.71			
	Within	348	366.45	1.05		

Note. Trials represent the different PBF prediction equations. Within subjects variance is the sum of error and trials variance and the result of a one-way ANOVA model with subjects effect F = 324.56 (female) and F = 206.58 (male). MS_S is mean squares for subjects variance. MS_T is mean squares for trials variance. MS_E is mean squares for error variance. MS_W is mean squares for within subject variance. Noteworthy that MS_E and MS_W terms are similar, hence ICCs are also similar.

Table 5. Intraclass correlation coefficients (ICCs) for percent body fat (PBF) measurements from five different prediction equations by sex

Sex	Model	ICC	LL	UL	F	*p*
Female	ICC(1,1)a	.985	.976	.991	324.563	<.0001
	ICC(2,1)a	.985	.968	.992	525.377	<.0001
	ICC(3,1)c	.991	.985	.994	525.377	<.0001
	ICC(2,1)c	.991	.985	.994	525.377	<.0001
	ICC(3,1)a	.985	.968	.992	525.377	<.0001
Male	ICC(1,1)a	.976	.968	.983	206.584	<.0001
	ICC(2,1)a	.976	.943	.988	419.922	<.0001
	ICC(3,1)c	.988	.984	.992	419.922	<.0001
	ICC(2,1)c	.988	.984	.992	419.922	<.0001
	ICC(3,1)a	.976	.943	.988	419.922	<.0001

Note. ICC(1,1) is a one-way random effects model, single measure. ICC(2,1) is a two-way random effects model, single measure. ICC(3,1) is a two-way mixed effects model, single measure. The 'a' subnotation on ICC refers to absolute agreement reliability. The 'c' subnotation on ICC refers to consistency reliability. The first 3 ICCs were defined by Strout and Fleiss and the last 2 ICCs defined by McGraw and Wong. ICC values for ICC(2,1)a and ICC(3,1)a are mathematically equivalent but differ in interpretation. ICC values for ICC(2,1)c and ICC(3,1)c are mathematically equivalent but differ in interpretation.
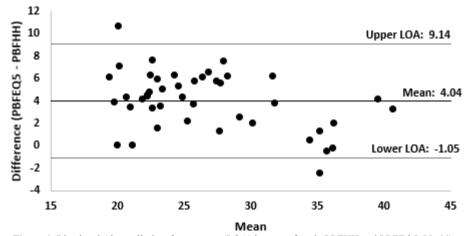
Figure 1. Bland and Altman limits of agreement (LOA) between female PBFHH and PBFEQ5 (N=44)
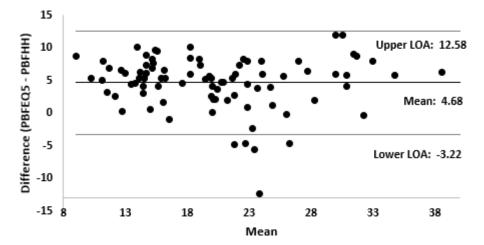

Figure 2. Bland and Altman limits of agreement (LOA) between male PBFHH and PBFEQ5 (N=87)

Table 6. Convergent validity evidence for percent body fat (PBF) measurements from five different prediction equations and PBF from bioelectric impedance (PBFHH) by sex

Sex	PBFEQ1	PBFEQ2	PBFEQ3	PBFEQ4	PBFEQ5
Female	.913	.917	.918	.917	.918
Male	.817	.819	.822	.822	.821

Note. All Pearson correlation coefficients were significant (ps < .0001).

Table 7. Equivalence reliability and convergent validity evidence for obesity classification from five different prediction equations and bioelectric impedance (BIA) by sex

Sex	Reliability			Validity		
	Kappa (κ)	LL	UL	Kappa (κ)	LL	UL
Female	.922	.837	1.007	.861	.674	1.000
Male	.756	.659	.853	.426	.213	.640

Note. Reliability Kappa statistics are Fleiss Kappa statistics for overall agreement across the five PBF prediction equation obesity classifications. Cutoff values for obese classification were ≥25% for male and ≥32% for female.

Finally, Table 7 displays equivalence reliability and convergent validity evidence for the prediction equations to classify participants into obese and non-obese categories by sex. The five PBF prediction equations' obesity classification showed substantial (κ = .756) agreement in male and almost perfect (κ = .922) agreement in female analyses. Additionally, obesity classification by PBFEQ5 successfully converged with BIA classification, showing

moderate (κ = .426) agreement in male and almost perfect (κ = .861) agreement in female analyses.

4. Discussion

The purpose of this study was to present psychometric evidence as to the extent to which prediction equations can accurately assess PBF and accurately classify individuals into obesity groups in a college student population. Since reliability is a prerequisite for validity, determining the equivalence of commonly used prediction equations was the first objective of this study. Using several ICC models, results consistently showed that BMI-based prediction equations have excellent agreement in estimating PBF. Reliability evidence was similar for both random effects models (i.e., generalizing to the population of all similar BMI-based PBF prediction equations) and mixed effects models (i.e., limiting the generalizations to only the five PBF prediction equations in this study). Similarly, in terms of obesity classification equivalence of the PBF prediction equations, results showed substantial to almost perfect agreement. Therefore, results from this study support the use of prediction equations to reliably measure PBF and reliably classify individuals into obesity groups. Furthermore, both estimation of PBF and obesity classification by prediction equation adequately converged with estimation and

classification by BIA, indicating that PBF prediction equations and BIA measure a similar construct. Albeit, a caveat worth noting, PBF measurements were systematically larger by prediction equation than by BIA. Therefore, future research may be needed to investigate which method (i.e., prediction equation vs. BIA) is responsible for this bias and/or conduct similar research using a criterion PBF method.

The primary limitation of this study is its use of a college student population. Therefore, the psychometric properties reported in this study should not be confusingly assumed for other populations. Another limitation worth mentioning is the use of BMI as an input variable in the PBF prediction equations. Specifically, this study did not assess the measurement error in height and weight measurement of study participants. Given this limitation, a different reliability study of BC assessment found BMI measurement to be reliable in the same study population via test-retest (stability) method [21].

5. Conclusions

Results from this study clearly support the use of BMI-based prediction equations to both measure PBF and classify individuals into obesity groups in college student populations. Implications for PBF prediction equation use are vast and include easier more practical body composition assessment for fitness educators, trainers, and researchers. However, further research may be warranted to determine the extent to which prediction equations overestimate PBF in this population.

Acknowledgements

References

[1]   Kenney, W.L., Wilmore, J.H. and Costill, D.L., 2015. Physiology of sport and exercise. Human kinetics.

[2]   Morrow Jr, J.R., Mood, D., Disch, J. and Kang, M., 2015. Measurement and Evaluation in Human Performance, 5E. Human Kinetics.

[3]   Mittal, R., Goyal, M.M., Dasude, R.C., Quazi, S.Z. and Basak, A., 2011. Measuring obesity: results are poles apart obtained by BMI and bio-electrical impedance analysis. Journal of Biomedical Science and Engineering, 4(11), p.677.

[4]   Kanellakis, S., Skoufas, E., Khudokonenko, V., Apostolidou, E., Gerakiti, L., Andrioti, M.C., Bountouvi, E. and Manios, Y., 2017. Development and validation of two equations based on anthropometry, estimating body fat for the Greek adult population. Obesity, 25(2), pp.408-416.

[5]   Deurenberg, P., Weststrate, J.A. and Seidell, J.C., 1991. Body mass index as a measure of body fatness: age-and sex-specific prediction formulas. British journal of nutrition, 65(2), pp.105-114.

[6]   Deurenberg, P., Yap, M. and Van Staveren, W.A., 1998. Body mass index and percent body fat: a meta-analysis among different ethnic groups. International journal of obesity, 22(12), p.1164.

[7]   Gallagher, D., Visser, M., Sepulveda, D., Pierson, R.N., Harris, T. and Heymsfield, S.B., 1996. How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups?. American journal of epidemiology, 143(3), pp.228-239.

[8]   Jackson, A.S., Pollock, M.L. and Ward, A.N.N., 1980. Generalized equations for predicting body density of women. Medicine and science in sports and exercise, 12(3), pp.175-181.

[9]   Jackson, A.S., Stanforth, P.R., Gagnon, J., Rankinen, T., Leon, A.S., Rao, D.C., Skinner, J.S., Bouchard, C. and Wilmore, J.H., 2002. The effect of sex, age and race on estimating percentage body fat from body mass index: The Heritage Family Study. International journal of obesity, 26(6), p.789.

[10]   American College of Sports Medicine. (2013). ACSM's guidelines for exercise testing and prescription. Lippincott Williams & Wilkins.

[11]   Omron Fat Loss Monitor. (2012). Model HBF-306C. Omron Healthcare Co., Ltd.

[12]   Shrout, P.E. and Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychological bulletin, 86(2), p.420.

[13]   McGraw, K.O. and Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychological methods, 1(1), p.30.

[14]   Bland, J.M. and Altman, D., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. The lancet, 327(8476), pp.307-310.

[15]   Guido, J.J., 2007. Guido's Guide to PROC FREQ–A Tutorial for Beginners Using the SAS® System. In Proceedings of the 20th annual North East SAS Users Group Conference, Baltimore, MD.

[16]   Chen, B., Zaebst, D. and Seel, L., 2005, April. A macro to calculate kappa statistics for categorizations by multiple raters. In Proceeding of the 30th Annual SAS Users Group International Conference (pp. 155-30).

[17]   Koo, T.K. and Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine, 15(2), pp.155-163.

[18]   Landis, J.R. and Koch, G.G., 1977. The measurement of observer agreement for categorical data. biometrics, pp.159-174.

[19]   Carlson, K.D. and Herdman, A.O., 2012. Understanding the impact of convergent validity on research results. Organizational Research Methods, 15(1), pp.17-32.

[20]   Cody, R. and Smith, J., 2005. Applied statistics and the SAS programming language. 403 p.

[21]   Hart, P. D. (2017). Test-retest stability of four common body composition assessments in college students. Journal of Physical Fitness, Medicine & Treatment in Sports. 1(2): 555561.