Original

Repeatability and Reproducibility of Anthropometric Measurements: An Evaluation of Intra- and Inter-Observer Reproducibility Among Students in the Faculty of Basic Medical Sciences, Ebonyi State University Abakaliki

[1]Arinze Francis Obasi, [1]Chiemeka Nwankwo Okoro, [1]Theresa Ebele Efor

[1]Department of Mathematics and Statistics, Faculty of Sciences, Ebonyi State University, Abakaliki, Nigeria

Corresponding author: Obasi Arinze Francis, Department of Mathematics and Statistics, Faculty of Sciences, Ebonyi State University, Abakaliki, Nigeria. arimatic89@gmail.com, +2348064439784

Article history: Received 13 January 2026, Reviewed 07 March 2026, Accepted for publication 22 March 2026

ABSTRACT

Background: Accurate anthropometric measurements are fundamental in clinical practice, sports science, and epidemiological research, yet their reliability can be influenced by methodological and sociocultural factors. This study investigated the reliability of anthropometric measurements with emphasis on gender-related influences in intra- and inter-rater reproducibility among ninety-six undergraduate students aged 18–30 years at Ebonyi State University, Abakaliki.

Methods: Standardized instruments were used to assess height, weight, head girth, neck girth, arm girth (relaxed and flexed), forearm girth, waist circumference, and gluteal girth, following ISAK protocols. Two trained raters, one male and one female, independently measured both male and female participants to evaluate same-gender and cross-gender reproducibility. Reliability was assessed using Intraclass Correlation Coefficient (ICC), Bland–Altman analysis, Cronbach's Alpha, and correlation coefficients.

Results: Results showed excellent reproducibility for height and weight across all groups (ICC > 0.99), while circumference-based parameters such as waist and gluteal girth demonstrated weaker agreement, particularly in cross-gender assessments (male gluteal girth ICC = 0.426; female gluteal girth ICC = 0.291). Same-gender measurements consistently yielded higher reliability, whereas opposite-gender measurements introduced greater variability, especially in culturally sensitive body regions.

Conclusion: The study concludes that although measurement errors cannot be completely eliminated, adherence to standardized protocols and incorporation of gender-sensitive approaches can substantially improve accuracy and reproducibility in anthropometric research.

Keywords: Anthropometric measurements, Gender-related influences, Reproducibility, Undergraduate students, Reliability, Interclass, Intraclass

How to cite this article

Obasi AF, Okoro CN, Efor TE. Repeatability and Reproducibility of Anthropometric Measurements: An Evaluation of Intra- and Inter-Observer Reproducibility among Students in the Faculty of Basic Medical Sciences, Ebonyi State University Abakaliki. The Nigerian Health Journal 2026; 26(1):309 – 319.
https://doi.org/10.71637/tnhj.v26i1.1302

INTRODUCTION

The measurement of anthropometric parameters remains a cornerstone of health assessments, sports performance evaluations, and ergonomic studies, and it is increasingly employed in school-based and community health research to evaluate body composition and growth trends.[1–3] Parameters such as height, mass, head girth, neck girth, arm girth (relaxed and flexed), forearm girth, wrist girth, waist circumference, and gluteal girth can be clinically assessed using standardized protocols and instruments.[4] Yet, despite the availability of guidelines, relatively few studies have examined how rater-related factors particularly gender affect the reproducibility of these measurements. Where such studies exist, they often report acceptable reliability but are limited by small sample sizes or by designs that focus on comparing measurement techniques rather than systematically evaluating reliability across gender.[5,6]

Reliability, in this context, refers to the degree to which repeated measurements of the same parameter yield consistent results.[5] Ensuring high reliability is essential to avoid misinterpreting variability as biological difference rather than methodological inconsistency. Reliability is typically assessed through intra-rater reliability, which examines the consistency of repeated assessments by the same rater, and inter-rater reliability, which evaluates the consistency between different raters measuring the same participant. When reliability is low, conclusions about population health trends may be distorted, undermining the validity of interventions and research outcomes. [6]

Gender dynamics introduce an additional layer of complexity. In many cultural contexts, particularly within sub-Saharan Africa, participants may experience discomfort when being measured by someone of the opposite sex, leading to subtle postural adjustments, hesitancy, or incomplete cooperation. Female participants measured by male raters may alter their posture due to modesty concerns, affecting waist or gluteal girth measurements, while male participants measured by female raters may unconsciously adjust muscle tension, influencing arm girth measurements. Beyond cultural sensitivities, gender-based differences in grip strength, tape placement, and hand positioning may also contribute to systematic discrepancies, thereby reducing reproducibility.

To address these challenges, this study employs multiple statistical tools to evaluate reproducibility with a focus on gender comparison. The Intraclass Correlation Coefficient (ICC) is used to assess agreement and consistency between raters or repeated measures, Bland-Altman analysis identifies systematic biases and inconsistencies, the Coefficient of Reliability (R) quantifies variance attributable to measurement error, and Cronbach Alpha evaluates internal consistency, indicating how closely related a set of measurements are as a group. [5,6] Together, these tools provide a comprehensive framework for examining whether gender contributes to systematic biases in anthropometric measurements and for comparing the effectiveness of different statistical approaches.

The aim of this study is therefore to evaluate the reliability of anthropometric measurements in an academic setting, with particular emphasis on gender-related differences in intra-rater and inter-rater reproducibility. By providing quantitative estimates of expected values when measurements are repeated by the same or different raters, this research seeks to determine whether gender contributes to systematic biases and to inform the development of more culturally sensitive and reliable measurement protocols.

MATERIALS AND METHODS

Study Design and Setting: The study employed an institutional-based cross-sectional design which involved the use of quantitative research methods. It was conducted among undergraduate students of the Faculty of Basic Medical Sciences at Ebonyi State University (EBSU), Abakaliki, during the 2024 – 2025 academic session. The Presco Campus, where the faculty is located, served as the study site because it provides a stable and diverse pool of students across different departments and levels. Participants were selected based on their willingness to participate in the study.

Study Population, Sample Size and Sampling Procedure: The study population comprised undergraduate students enrolled in the Faculty of Basic Medical Sciences at Ebonyi State University (EBSU), Abakaliki, during the 2024-2025 academic session. According to the official Faculty Register, the estimated student population was approximately 1,200.

The minimum sample size was determined using the standard formula for cross-sectional studies:

$$Sample\ size\ (n) = \frac{N}{1+N(e^2)}$$

This formula gives the minimum number of samples with two-sided type I error probability alpha (α). The usual choice for alpha (α) is 5%

Print ISSN: 0189-9287 Online ISSN: 2992-345X

Where n = sample size for the study.

$e = 0.1 =$ error terms.

$Where\ N = 1200;\ e = 0.10$

$= \frac{1200}{1+1200(0.1^2)}$

$= \frac{1200}{1+1200(0.01)}$

$= \frac{1200}{1+1200(0.01)}$

$= \frac{1200}{1+12}$

$= \frac{1200}{13}$

$n = 92.308 \cong 92$

Therefore, the minimum sample size for this study is = 92 subjects

The sample size was adjusted for the attrition rate of 10% = 9.231 ≅ 9 from the minimum sample size of 92 to 102 subjects.

A random sampling procedure was initially employed at the faculty level to select one out of the three departments (Anatomy, Physiology, and Biochemistry). The Department of Anatomy was selected. Within Anatomy, students were stratified by gender (male and female), and proportionate random sampling was applied to ensure balanced representation (see Figure 1). Eligible students were approached during school hours and invited to participate voluntarily. Although the calculated sample size was 102, the final recruited sample comprised 96 participants (35 males and 61 females).

Bland-Altman Plot and Analysis: The Bland-Altman (mean-difference or limits of agreement) plot and analysis is used to compare two measurements of the same variable. That is, it is a method comparison technique. For example, an expensive measurement system might be compared with a less expensive one or an intrusive measurement system might be compared to one that is less intrusive. The technique is documented in previous reports by Martin and Althman. [7]

Bias: The bias between the two tests is measured by the mean of the differences calculated in the usual fashion as $\bar{d} = \frac{1}{n}\sum_{k=1}^{n} d_k$

Limits of Agreement: Limits of agreement between the two tests are defined by a 95% prediction interval of a particular value of the difference which are computed as follows $\bar{d} \pm 1.96 S_d$

Where:

$$S_d = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(d_k - \bar{d}\right)^2}$$

Martin and Althman provided the following variances and confidence intervals for the bias and the limits of agreement, assuming that the differences are normally distributed.[7]

The linear Correlation Coefficient: The correlation coefficient can be computed from either of the results

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} \quad (1)$$

Or $$r^2 = \frac{\sum(y_{est}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{explained\ variation}{total\ variation} \quad (2)$$

which for linear regression are equivalent. The formula (2) is often referred to as the product-moment formula for linear correlation.

Formulas equivalent to those above, which are often used in practice, are

$$r = \frac{n\sum xy - (\sum x)(y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (3)$$

And $$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{(\bar{x}^2 - \bar{x}^2)(y^2 - y^2}} \quad (4)$$

Cronbach Alpha: Cronbach's Alpha is a statistical tool used to evaluate the internal consistency and reliability of measurement scales or repeated assessments. In the context of anthropometric measurements, it provides an estimate of how closely related repeated measurements of the same parameter are, thereby indicating the stability of the measurement process. The coefficient ranges from 0 to 1, with higher values reflecting stronger reliability. A Cronbach Alpha value above 0.70 is generally considered acceptable, values above 0.80 indicate good reliability, and values above 0.90 suggest excellent consistency.

Intraclass Correlation Coefficient (ICC): The Intraclass Correlation Coefficient (ICC) is a statistical measure used to assess the reliability of measurements when multiple raters or repeated observations are involved. It quantifies the degree of agreement among measurements taken under similar conditions. The mathematical models for ICC are derived from Analysis of Variance (ANOVA) and depend on the type of reliability being assessed.

ICC is generally expressed as the ratio of true variance to total variance, which can be formulated as:

$$ICC = \frac{\sigma^2_{subject}}{\sigma^2_{subject} + \sigma^2_{error}}$$

where:

- $\sigma^2_{subject}$ is the variance due to differences between subjects.

Print ISSN: 0189-9287 Online ISSN: 2992-345X

σ^2_{error} is the variance due to measurement error.

Method of Data Collection and Data Analysis: Anthropometric assessments were conducted using standardized instruments: a stadiometer for height (Prestige) [8], a digital weighing scale for body mass (Seca Digital Scale) [9], and a non-elastic measuring tape for girth and circumference measurements (Seca). [10] All instruments were calibrated at the beginning of each session to maintain precision. Measurements were taken with participants barefoot, dressed in light clothing, and positioned according to standardized postures outlined by the International Society for the Advancement of Kinanthropometry (ISAK) guidelines. [11,12]

Informed consent was obtained from all participants prior to data collection. Ethical approval for the study was granted by the Department, and all procedures adhered strictly to established ethical principles for research involving human participants. Confidentiality and voluntary participation were ensured throughout the study. Two trained raters independently measured each participant in a randomized sequence. Each rater assessed both male and female participants. Assessments were conducted within 3–4 minutes of each other to minimize the influence of posture changes or fatigue. Each measurement was documented by an assistant, and raters were blinded to all previous results to reduce bias. Intra-rater reproducibility was determined by comparing two consecutive measurements taken by the same rater for each participant, while inter-rater reproducibility was assessed by analyzing the two independent series of measurements obtained by both raters across the participants. Rater A is a trained male research assistant and rater B is a trained female research assistant. This design resulted in four examinations per parameter. Reliability was evaluated using the Intraclass Correlation Coefficient (ICC), Bland–Altman analysis, Coefficient of Reliability (R), and Cronbach's Alpha. ICC values were calculated for each test–retest evaluation (e.g., 1st rater A1–A2, 2nd rater B1–B2 for intra-rater assessment; first- and second-rater comparisons for inter-rater assessment), after which mean ICCs and their ranges were reported.[13]

The Statistical Package for Social Sciences (IBM SPSS) version 30 was used to perform analysis, while Microsoft Office Excel 365 was used for data collation, computations and drawing of charts. The expected range of variability between repeated evaluations was estimated using paired score differences (test minus retest). Scatter plots were constructed by plotting each test–retest difference against its mean, and the mean value with its standard deviation (SD) was calculated to describe the spread of differences. The 95% confidence intervals (CI) of the mean difference were considered the boundaries of expected variability. Because mean differences between repeated measurements tended to be greater than zero, this value was incorporated into the standard calculation of 95% CI, which were also expressed as percentage values around the mean measurement.[7,14]

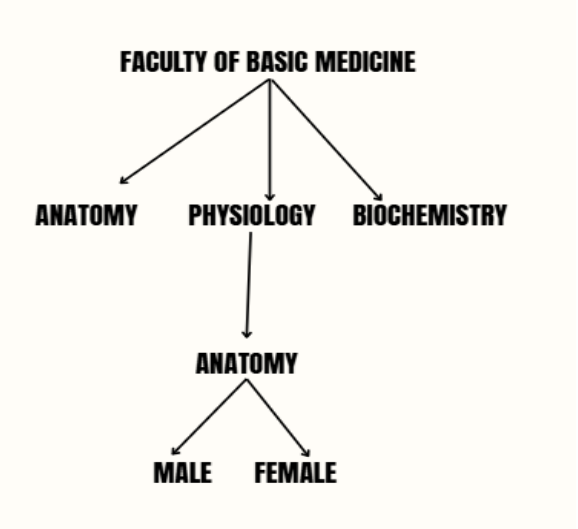

Figure1. Multi stage sampling procedures of the sample section

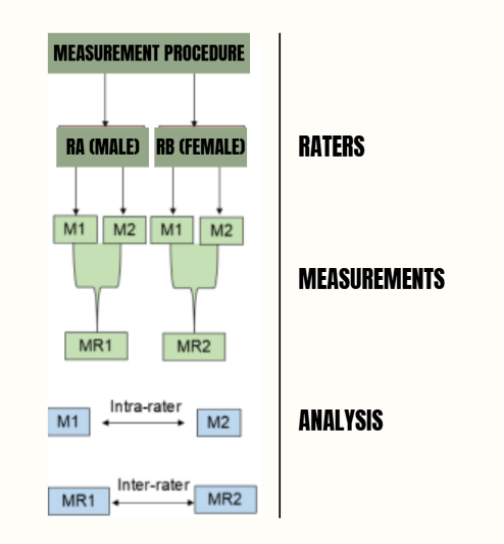

Figure 2: Protocol of the Reproducibility Study

Print ISSN: 0189-9287 Online ISSN: 2992-345X

Abbreviations: RA, Rater A; RB, Rater B; M1, first measurement; M2, repeat measurement; MR1, average of repeated measurements by rater 1; MR2, average of repeated measurements by rater 2

RESULT

The results of the data analysis are presented in this chapter along with their interpretations. There was 100% response rate as participants who refused to participate/give consent got replaced by other students, hence the sample used was 96. The demographic data of the 96 students is shown in Figures 3 & 4. Most respondents fall within the 20–24 age range 61 (63.5%), followed by those under 20 years 28 (29.2%). A smaller proportion are in the 25–29 bracket 7 (7.3%). Females make up the majority 61 (63.5%) compared to males 35 (36.5%).

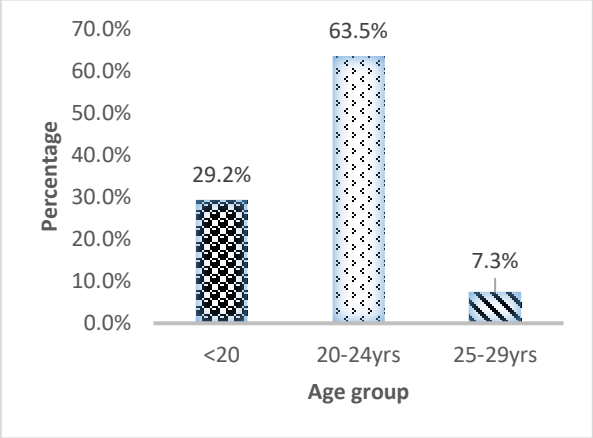

Figure 3: Bar chart showing age group of the students

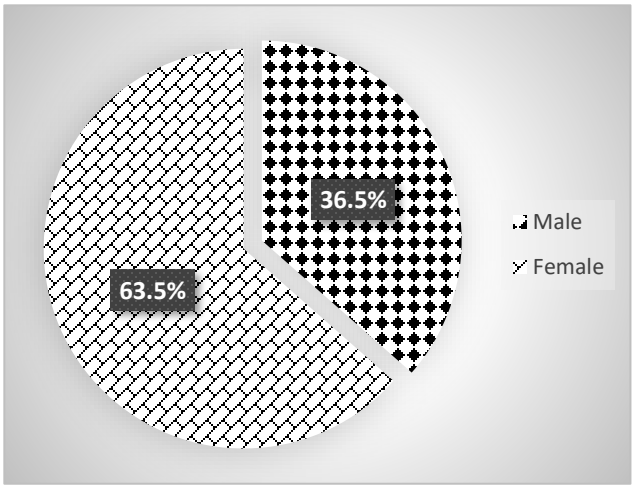

Figure 4: Bar chart showing gender of the students

Print ISSN: 0189-9287 Online ISSN: 2992-345X

Table 1: Descriptive statistics of anthropometric parameters obtained by each of the observers

Variable	Intra-observer (Observer A)			Intra-observer (Observer B)		
	A1	A2		B1	B2	t (P-value)
	Mean ± SD	Mean ± SD	t (P-value)	Mean ± SD	Mean ± SD	
Height	167.17±6.05	167.18±6.03	-0.193 (0.848)	167.33±6.26	167.36±6.21	-0.515 (0.608)
Weight	61.75±7.83	61.86±7.92	-1.014 (0.313)	62.19±7.72	61.87±7.58	1.489 (0.14)
Head girth	56.32±1.92	56.16±2.19	1.387 (0.169)	56.22±1.98	56±1.81	2.161 (0.033)*
Neck girth	33.51±2.9	33.18±2.93	3.126 (0.002)*	33.57±2.95	33.55±3.06	0.204 (0.839)
Girth relax	26.88±2.15	26.85±2.19	0.341 (0.734)	26.78±2.34	26.76±2.61	0.169 (0.866)
Arm flex	29.54±2.58	28.66±3.86	2.791 (0.006)*	29.05±2.79	28.92±2.49	0.745 (0.458)
Forearm	25.7±1.85	25.65±1.84	0.473 (0.637)	25.63±1.94	25.64±1.82	-0.178 (0.859)
Waist circumference	70.66±7.02	71.94±5.65	-2.156 (0.034)*	70.47±6.03	71.37±5.72	-1.786 (0.077)
Gluteal girth	92.34±6.13	92.06±6.07	0.435 (0.665)	91.82±6.87	90.72±5.92	1.609 (0.111)

**Statistically Significant*

The result in table 1 shows the descriptive statistics of anthropometric parameters obtained by each of the observers. The anthropometric parameters measured by both observers generally showed consistency across repeated trials, with most variables recording no significant differences between the first and second measurements. For height, weight, forearm, girth relax, and gluteal girth, both observers demonstrated reliable measurements as indicated by non-significant p-values. However, some inconsistencies were observed. Observer A recorded significant differences in neck girth (p = 0.002), arm flex (p = 0.006), and waist circumference (p = 0.034). Observer B showed inconsistency in head girth (p = 0.033), while other parameters remained stable.

Table 2: Intra-observer agreement of anthropometric parameters of male participants obtained by each of the observers (n = 35)

Variable	Intra-observer (Observer A)				Intra-observer (Observer B)			
	ICC (95% CI)	**BA (95% LOA)**	**Cronbach Alpha**	**R**	**ICC (95% CI)**	**BA (95% LOA)**	**Cronbach Alpha**	**R**
Height	.996 (0.98 - 1)	0.03 (-1.19 - 2.95)	0.996	0.993	.999 (0.99 - 1)	-0.03 (-1.26 - 3.1)	0.999	0.998
Weight	.999 (1 - 1)	0.11 (-0.93 - 2.35)	0.999	0.998	.942 (0.79 - 0.94)	0.33 (-3.9 - 9.8)	0.942	0.891
Head girth	.989 (0.96 - 0.99)	-0.05 (-0.89 - 2.17)	0.989	0.983	.966 (0.87 - 0.97)	0.21 (-1.69 - 4.28)	0.966	0.935
Neck girth	.979 (0.92 - 0.98)	0.16 (-1.88 - 4.72)	0.979	0.959	.980 (0.92 - 0.98)	0.02 (-1.74 - 4.31)	0.980	0.962
Girth relax	.973 (0.9 - 0.97)	-0.26 (-1.57 - 3.75)	0.973	0.948	.754 (0.55-0.85)	0.02 (-1.86 - 4.61)	0.754	0.817
Arm flex	.986 (0.95 - 0.99)	0.17 (-1.01 - 2.58)	0.986	0.972	.964 (0.87 - 0.96)	0.12 (-3.07 - 7.65)	0.964	0.939
Forearm	.984 (0.94 - 0.98)	0.03 (-0.56 - 1.4)	0.984	0.972	.991 (0.96 - 0.99)	-0.01 (-1.36 - 3.36)	0.991	0.982
Waist circumference	.926 (0.74 - 0.93)	0.18 (-3.99 - 9.95)	0.926	0.922	.644 (0.17 - 0.7)	-0.89 (- 10.49 - 25.46)	0.644	0.492
Gluteal girth	.576 (0.09 - 0.65)	0.48 (-10.32 - 25.74)	0.576	0.409	.426 (0.20-0.62)	1.1 (-12.03 - 30.28)	0.426	0.376

Abbreviation: ICC = Inter Class Correlation; CI = Confidence interval; BA = Bland-Altman; 95% LOA = 95% Limit of Agreement; R = Correlation

The result in Table 2 presents the intra-observer agreement of anthropometric parameters of male participants obtained by each observer. Most measurements demonstrated high reliability, with strong intra-class correlation coefficients (ICC),

Print ISSN: 0189-9287 Online ISSN: 2992-345X

Cronbach Alpha values, and correlation coefficients (R), indicating consistency across repeated measurements. For Observer A, height, weight, head girth, neck girth, arm flex, and forearm all showed excellent agreement (ICC > 0.97), while waist circumference recorded good reliability (ICC = 0.926). However, gluteal girth showed weaker agreement (ICC = 0.576), suggesting variability in repeated measurement. For Observer B, height also demonstrated excellent agreement (ICC = 0.999), but weight (ICC = 0.942), head girth (ICC = 0.966), neck girth (ICC = 0.980), arm flex (ICC = 0.964), and forearm (ICC = 0.991) showed slightly lower but still strong reliability compared to Observer A. However, girth relax (ICC = 0.754), waist circumference (ICC = 0.644), and gluteal girth (ICC = 0.426) revealed moderate to poor agreement, reflecting inconsistencies in repeated measurements by Observer B.

Table 3: Intra-observer agreement of anthropometric parameters of female participants obtained by each of the observers (n=61**)**

Variable	Intra-observer (Observer A)				Intra-observer (Observer B)			
	ICC (95% IC)	BA (95% LOA)	Cron bach Alpha	R	ICC (95% CI)	BA (95% LOA)	Cron bach Alpha	R
Height	.996 (0.99 - 0.99)	-0.04 (-1.14 - 2.79)	0.996	0.991	.993 (0.98 - 0.99)	-0.08 (-1.55 - 3.79)	0.993	0.987
Weight	.993 (0.98 - 0.99)	-0.24 (-2.69 - 6.52)	0.993	0.986	.997 (0.99 - 1)	-0.14 (-1.75 - 4.25)	0.997	0.994
Head girth	.878 (0.66 - 0.86)	0.28 (-2.41 - 6.09)	0.878	0.791	.900 (0.71 - 0.89)	0.3 (-1.82 - 4.65)	0.900	0.828
Neck girth	.953 (0.86 - 0.95)	0.43 (-1.61 - 4.2)	0.953	0.911	.973 (0.91 - 0.97)	-0.05 (-1.56 - 3.83)	0.973	0.948
Girth relax	.918 (0.76 - 0.91)	0.2 (-2.03 - 5.12)	0.918	0.848	.960 (0.88 - 0.95)	0.13 (-1.83 - 4.59)	0.960	0.928
Arm flex	.505 (0.1 - 0.54)	1.29 (-6.16 - 15.87)	0.505	0.390	.838 (0.65-0.89)	0.2 (-3.64 - 9.09)	0.838	0.731
Forearm	.860 (0.62 - 0.85)	0.05 (-2.2 - 5.46)	0.860	0.755	.924 (0.77 - 0.91)	-0.08 (-1.73 - 4.23)	0.924	0.860
Waist Circum-ference	.655 (0.27 - 0.66)	-2.12 (-15.9 - 38.19)	0.655	0.492	.795 (0.49 - 0.78)	-0.46 (-10.61 - 25.98)	0.795	0.661
Gluteal girth	**.504 (0.09 - 0.54)**	**0.16 (-13.13 - 32.51)**	**0.504**	**0.337**	**.790 (0.50-0.82)**	**2 (-10.39 - 26.68)**	**0.790**	**0.422**

Abbreviation: ICC = Inter Class Correlation; CI = Confidence interval; BA = Bland-Altman; 95% LOA = 95% Limit of Agreement; R = Correlation

The result in Table 3 presents the intra-observer agreement of anthropometric parameters of female participants obtained by each of the observers. Height and weight demonstrated excellent reliability for both observers, with ICC values above 0.99, Cronbach Alpha values close to 1.0, and narrow Bland–Altman limits of agreement, indicating highly consistent repeated measurements. Head girth and neck girth also showed strong agreement, with ICC values ranging from 0.878 to 0.973 and Cronbach Alpha values above 0.79, reflecting good reproducibility. Girth relax and forearm measurements recorded similarly high reliability, with ICC values above 0.86 and Cronbach Alpha values above 0.84, confirming stable intra-observer consistency. However, arm flex showed poor reliability for Observer A, with an ICC of 0.505 and Cronbach Alpha of 0.390, while Observer B achieved moderate agreement with an ICC of 0.838. Waist circumference and gluteal girth demonstrated weaker consistency, particularly for Observer A, with ICC values of 0.655 and 0.504 respectively, wider Bland–Altman limits, and lower correlation coefficients, suggesting substantial variability in repeated measures. Observer B performed somewhat better in these parameters, with ICC values of 0.795 for waist circumference and 0.790 for gluteal girth, though reliability remained moderate.

Table 4: Inter-observer agreement of anthropometric parameters of male participants obtained between the two observers (A & B) (n = 35)

Variable	ICC (95% CI)	BA (95% LOA)	Cronbach Alpha	R
Height	.997 (0.99 - 1)	-0.17 (-1.64 - 3.96)	0.997	0.995
Weight	.972 (0.89 - 0.97)	-0.23 (-3.58 - 8.73)	0.972	0.954
Head girth	.978 (0.92 - 0.98)	0.13 (-1.77 - 4.44)	0.978	0.958
Neck girth	.983 (0.94 - 0.98)	-0.22 (-1.77 - 4.26)	0.983	0.970

Print ISSN: 0189-9287 Online ISSN: 2992-345X

Variable	ICC (95% CI)	BA (95% LOA)	Cronbach Alpha	R
Girth relax	.775 (0.91 - 0.98)	0.1 (-1.82 - 4.55)	0.775	0.852
Arm flex	.958 (0.85 - 0.96)	0.11 (-3.32 - 8.26)	0.958	0.944
Forearm	.973 (0.9 - 0.97)	0.04 (-0.92 - 2.29)	0.973	0.948
Waist	.727 (0.75 - 0.93)	0.38 (-5.46 - 13.68)	0.727	0.866
Gluteal girth	.513 (0.02 - 0.61)	0.93 (-7.24 - 18.36)	0.513	0.366

The result in Table 4 presents the inter-observer agreement of anthropometric parameters of male participants obtained between the two observers. Height, weight, head girth, neck girth, forearm, and arm flex demonstrated excellent agreement, with ICC values above 0.95, Cronbach Alpha values close to 1.0, and narrow Bland–Altman limits of agreement, indicating highly consistent measurements between observers. Girth relax showed moderate reliability, with an ICC of 0.775 and Cronbach Alpha of 0.852, suggesting acceptable but less robust consistency. Waist circumference recorded lower agreement, with an ICC of 0.727 and wider limits of agreement, reflecting moderate variability between observers. Gluteal girth demonstrated the weakest reliability, with an ICC of 0.513 and a low correlation coefficient, indicating substantial measurement variability, as shown in Figure 12, where many of the disagreements were above the upper and lower agreement limits.

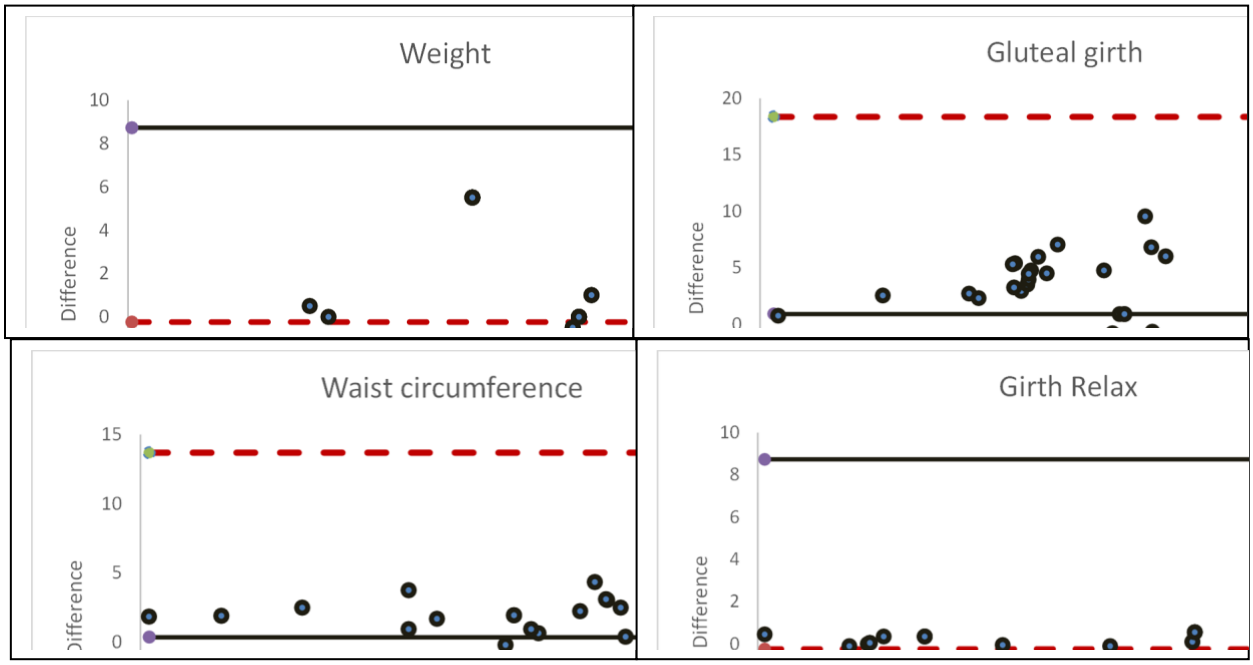

Table 5: Reliability Indices of the Intra Observer (ICC, Cronbach Alpha, and Correlation) of Anthropometric Parameters

Parameter	ICC	Cronbach Alpha	R
Height	0.998	0.997	0.995
Weight	0.995	0.997	0.991
Head girth	0.918	0.935	0.856
Neck girth	0.967	0.972	0.936
Girth relax	0.942	0.934	0.891
Arm flex	0.716	0.975	0.603

Print ISSN: 0189-9287 Online ISSN: 2992-345X

Parameter	ICC	Cronbach Alpha	R
Forearm	0.932	0.969	0.873
Waist circumference	0.734	0.934	0.593
Gluteal girth	0.633	0.996	0.464

Table 6: Descriptive Statistics of Reliability Measures (ICC, Cronbach Alpha, and R**)**

Groups	*Count*	*Sum*	*Average*	*Variance*
ICC	9	7.835	0.871	0.019
Cronbach Alpha	9	8.709	0.968	0.001
R	9	7.202	0.800	0.038

Table 7: Analysis of Variance (ANOVA) of Reliability Measures

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.127	2	0.064	3.310	0.054	3.403
Within Groups	0.461	24	0.019			
Total	0.589	26				

The effectiveness of selected statistical methods in evaluating and comparing measurement reliability was demonstrated through the use of intraclass correlation coefficients (ICC), Cronbach Alpha, and correlation values. The results showed that there is no statistically significant difference between the three methods ($p = 0.054$), though the near-significance suggests subtle differences in sensitivity. Taken together, these findings indicate that ICC, Cronbach Alpha, and correlation are effective tools for evaluating measurement reliability.

DISCUSSION

This study investigated the reliability of anthropometric measurements with particular attention to gender-related influences on intra- and inter-rater reproducibility. The findings demonstrate that gender plays a significant role in the consistency of repeated assessments. Among male participants, Observer A recorded variability in relaxed arm girth, while Observer B showed inconsistencies in weight, forearm, and waist circumference. In female participants, weaker reproducibility was observed in parameters such as arm flex (ICC = 0.505 for Observer A), waist circumference (ICC = 0.655), and gluteal girth (ICC = 0.504). Observer B achieved moderately better agreement for these same parameters, with ICC values of 0.795 for waist circumference and 0.790 for gluteal girth. These findings are consistent with results of Njoku *et al.* [15], who reported that measurement errors were more pronounced when raters assessed individuals of the opposite gender, underscoring the importance of gender sensitivity in anthropometric reproducibility.

Height and weight demonstrated the highest reproducibility across both genders, with ICC values consistently above 0.99 and Cronbach Alpha values greater than 0.99. These parameters appear to be the most stable and least influenced by rater or participant gender. This observation aligns with [16], who highlighted that waist circumference and waist-to-height ratio were more predictive in women, while neck circumference and BMI were more relevant in men. In the present study, neck girth among males showed strong agreement (ICC = 0.979 for Observer A and ICC = 0.980 for Observer B), whereas female neck girth demonstrated slightly lower but still acceptable reliability (ICC = 0.953 and 0.973). These subtle differences suggest that gender influences not only the accuracy of measurements but also the reproducibility of repeated assessments, echoing [17], who identified gender differences in ultrasound reliability for adipose tissue.

Inter-rater comparisons revealed further gender-related discrepancies. Male participants experienced significant variability in waist circumference (ICC = 0.727) and gluteal girth (ICC = 0.513), while female participants

Print ISSN: 0189-9287 Online ISSN: 2992-345X

showed even weaker agreement for these same parameters (ICC = 0.406 for waist circumference and ICC = 0.291 for gluteal girth). These values highlight that circumference-based measures, particularly those involving sensitive body regions, are more prone to error when measured across raters of different genders. Conversely, parameters such as height and weight remained highly consistent across raters for both sexes (ICC = 0.997 for male height, ICC = 0.990 for female height; ICC = 0.972 for male weight, ICC = 0.994 for female weight). This pattern reflects the findings of [15], who reported that opposite-gender assessments introduced greater error in girth-related parameters. The weaker agreement in waist and gluteal girth among males also parallels [18], who found lower reproducibility in circumference measures, suggesting that these sites are inherently more challenging to measure consistently.

The use of multiple statistical approaches—Intraclass Correlation Coefficient (ICC), Cronbach's Alpha, correlation coefficients, and Bland–Altman plots—provided a comprehensive evaluation of measurement reliability. ICC and Cronbach's Alpha consistently demonstrated high reliability for stable parameters such as height (ICC = 0.998, Cronbach Alpha = 0.997) and weight (ICC = 0.995, Cronbach Alpha = 0.997). Correlation coefficients offered complementary insights, with values of R = 0.995 for height and R = 0.991 for weight, confirming strong reproducibility. Bland–Altman plots revealed systematic biases in sensitive parameters, particularly waist and gluteal girth, where limits of agreement were wide and differences exceeded ±10 units in some cases. These findings align with [17], who emphasized the diagnostic value of graphical tools in gage repeatability and reproducibility analysis, and [19], who advocated for ANOVA and Gage R&R techniques to identify sources of measurement variability. [20] similarly highlighted the importance of reliability coefficients in quantifying random error, particularly in educational assessments.

CONCLUSION

Although errors arising from the extraction of anthropometric data cannot be completely eliminated, they can be drastically reduced if standardized protocols are strictly followed. Training of measurers or anthropometrists has long been recognized as an effective means of minimizing error, particularly when guided by approaches such as ISAK. However, gender sensitivity remains a notable challenge to reproducibility, as males tend to be more accurate and consistent when measuring male participants, while females demonstrate greater reproducibility when measuring female participants. Cross-gender assessments, especially in sensitive regions such as the waist and gluteal girth, introduced higher variability and reduced reliability. This challenge can be mitigated by providing participants with durational training and orientation, enabling them to readily give consent, become accustomed to measurement protocols, overcome fear, and maintain confidence throughout the measuring sessions. By combining standardized procedures with gender-sensitive approaches, anthropometric research can achieve greater accuracy, reproducibility, and cultural appropriateness.

Declarations

***Ethical Approval*:** Not applicable

***Consent to Participate*:** Obtained

Competing Interest: The authors declared that there is no competing interest of any sort

***Funding*:** There is no external funding received by the authors

***Conflicts of Interest*:** The authors declared that there is no conflict of interest in this article

Use of AI Technology: The authors utilized AI tools such as Grammarly for grammatical checks

Acknowledgement: The author appreciates the cooperation of the respondents and the support of the Department of Mathematics and Statistics, Faculty of Sciences, Ebonyi State University, Abakaliki, Nigeria, during the conduct of this study

REFERENCES

1. Wang M, Song Y, Zhao X, Wang Y, Zhang M. Utilizing Anthropometric Measurements and 3D Scanning for Health Assessment in Clinical Practice. Physical Activity and Health. 2024;8(1):182–96.
2. Kobel S, Kirsten J, Kelso A. Anthropometry – Assessment of Body Composition. Dtsch Z Sportmed. 2022;73(3):106–11.
3. Casadei K, Kiel J. Anthropometric Measurement. National Library of Medicine [Internet]. 2022 Sep 26 [cited 2025 Dec 3];1–6. Available from: https://www.ncbi.nlm.nih.gov/books/NBK537315/
4. Kim M, Qiu X, Wang Y (Arthur). Interrater agreement in genre analysis: A methodological review and a comparison of three measures. Research Methods in Applied Linguistics. 2024 Apr 1;3(1):100097.

Print ISSN: 0189-9287 Online ISSN: 2992-345X

5. Nel S, de Man J, van den Berg L, Wenhold FAM. Statistical assessment of reliability of anthropometric measurements in the multi-site South African National Dietary Intake Survey 2022. Eur J Clin Nutr [Internet]. 2024 Nov 1 [ited 2025 Dec 3];78(11):1005. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11537951
6. Perumal N, Namaste S, Qamar H, Aimone A, Bassani DG, Roth DE. Anthropometric data quality assessment in multisurvey studies of child growth. American Journal of Clinical Nutrition. 2020 Jul 21;112:806S-815S.
7. Martin Bland J, Altman Dg. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. 1986 Feb 8;327(8476):307–10.
8. Warrier V, Krishan K, Shedge R, Kanchan T. Height Assessment. StatPearls [Internet]. 2023 Jul 25 [cited 2025 Dec 3]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK551524/
9. Bialocerkowski A. Measurement error and reliability testing: Application to rehabilitation. Int J Ther Rehabil [Internet]. 2008 Oct 1 [cited 2025 Dec 3]; Available from: https://www.academia.edu/125434029/Measurement_error_and_reliability_testing_Application_to_rehabilitation
10. Heymsfield SB, Bourgeois B, Ng BK, Sommer MJ, Li X, Shepherd JA. Digital Anthropometry: A Critical Review. Eur J Clin Nutr [Internet]. 2018 May 1 [cited 2025 Dec 3];72(5):680. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6411053/
11. Gibson RS. Anthropometric assessment of body composition. Principles Of Nutritional Assessment. 2023 Nov 2;273–98.
12. da Silva VS, Vieira MFS. International society for the advancement of kinanthropometry (Isak) global: International accreditation scheme of the competent anthropometrist. Revista Brasileira de Cineantropometria e Desempenho Humano. 2020;22:1–6.
13. Leah DO, Omokwa EA, Yakubu SI, Adeyinka NOF, Florence O. Influence of Occupational Physical Activity on Anthropometric Profile and Body Composition of Bricklayers in Kwara state, Nigeria. Exercise Medicine. 2018 Apr 10; 2:7.
14. Schober P, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. Anesth Analg. 2018 May 1;126(5):1763–8.
15. Njoku C, Oa N, Oa E, Ad S. Impact of Gender Sensitivity on Anthropometric Measurements. Vol. 14, Impact of Gender Sensitivity on Anthropometric Measurements the Journal of Anatomical Sciences. 2023.
16. Brambilla P, Bedogni G, Heo M, Pietrobelli A. Waist circumference-to-height ratio predicts adiposity better than body mass index in children and adolescents. Int J Obes. 2013 Jul;37(7):943–6.
17. Simkus A, Coolen-Maturi T, Coolen FPA, Bendtsen C. Statistical Perspectives on Reproducibility: Definitions and Challenges. J Stat Theory Pract. 2025 Sep 1;19(3).
19. Zanobini A, Sereni B, Catelani M, Ciani L. Repeatability and Reproducibility techniques for the analysis of measurement systems. Measurement (Lond). 2016 May 1;86:125–32.
20. Downing SM. Reliability: on the reproducibility of assessment data. Med Educ [Internet]. 2004 Sep [cited 2025 Dec 3];38(9):1006–12. Available from: https://pubmed.ncbi.nlm.nih.gov/15327684/
21. Larson-Meyer DE, Woolf K, Burke L. Assessment of nutrient status in athletes and the need for supplementation. Vol. 28, International Journal of Sport Nutrition and Exercise Metabolism. Human Kinetics Publishers Inc.; 2018. p. 139–58.

Print ISSN: 0189-9287 Online ISSN: 2992-345X