International Journal of Enhanced Research in Medicines & Dental Care (IJERMDC),
ISSN: 2349-1590, Vol. 12 Issue 11, November 2025

Assessing Agreement in Measurement of Orthodontic study Models: Digital Caliper VS Manual (Dividers and scales)

Dr. Rhutuja Rahul Kate[1], Dr. Pratap. N. Mane[2], Dr. Aarya Pratap Mane[3]

[1,3]BDS (Department of Orthodontics and Dentofacial Orthopedics, SDS, KVV, Karad)
[2]MDS (Assistant Professor, Department of Orthodontics and Dentofacial Orthopedics, SDS, KVV, Karad)

ABSTRACT

Introduction: Accurate measurement of orthodontic study models is crucial for diagnosis, treatment planning, and case documentation. While manual methods using dividers and scales have long been the standard, digital calipers offer potential advantages in precision and repeatability. This study aims to assess the agreement and reliability between these two measurement techniques.

Materials and Methods: A total of 20 orthodontic study models were measured by two calibrated operators using both a digital caliper and manual dividers with scales. Measurements included mesiodistal widths, buccolingual widths, facial axis of the clinical crown (FACC), and various arch dimensions. Data were analysed using repeated-measures ANOVA, MANOVA, to evaluate consistency, error variance, and method-related effects.

Results: Digital calipers demonstrated higher measurement precision, with lower error variance (0.012–0.020 mm) compared to manual methods (0.014–0.056 mm). Reliability coefficients for both methods approached 1.0, indicating excellent reliability. Statistically significant differences were found between methods ($p < 0.05$) with digital caliper and manual methods.

Conclusion: Both digital and manual methods are reliable for orthodontic model measurements. However, digital calipers offer enhanced precision, reduced observer bias, and better repeatability, making them suitable for more detailed assessments. Clinicians should select measurement tools based on case complexity, accuracy requirements, and available resources.

INTRODUCTION:

In orthodontics, diagnostic aids are integral part of treatment planning, of which research models are one of the most relevant.[2] These models are essential for treatment planning, progress evaluation and case documentation. Accurate measurements are essential for ensuring proper patient care.[8]

Commonly used instruments include dividers, rulers, Vernier calipers, and 3D digital scanners.[9] Despite growing use of digital tools in orthodontics, concerns remain regarding their consistency and reliability compared to traditional manual methods.[10]

Digital calipers offer advantages including fast readings, direct numerical output, and reduced observer bias [11]. Manual methods, while considered the gold standard due to simplicity and affordability, are operator-dependent and prone to errors, especially in complex cases.[12]

Despite the growing adoption of digital tools in orthodontics, questions remain regarding the consistency and reliability of these instruments compared to manual methods. Discrepancies in measurements can influence clinical decisions, highlighting the need for systematic comparison and validation of measurement tools.

This comparison focuses on two common measurement tools: digital caliper and manual dividers with scales. Traditionally, manual methods using dividers and scales have been the gold standard in clinical and research settings due to their simplicity and cost-effectiveness. However, with the advent of digital technology, digital calipers have gained popularity for their ease of use, speed , direct numerical readouts, consistent and more standardize measurements, minimizing observer bias and enhancing accuracy.

This research aims to assess the agreement between digital calipers and traditional manual tools (dividers and scales) in measuring orthodontic study models. Understanding these differences will aid in determining whether digital tools can be a reliable substitute for manual methods or if traditional techniques still hold critical value in clinical orthodontics.

MATERIALS AND METHODS

Ethical approval was obtained from the Institutional Ethics Committee, Protocol Number 092/2025-2026. Orthodontic plaster study models were sourced from the Department of Orthodontics and Dentofacial Orthopedics records. Inclusion criteria required fully erupted permanent teeth extending from the right first molar to the left first molar, with a displacement of less than 4 mm and no crown anomalies[1]. Exclusion criteria included broken or poorly formed casts.

Two trained examiners (Operator 1 and Operator 2) independently measured each model twice, one week apart, after calibration on 20 random samples to ensure measurement consistency[13]. Calibration involved repeated measurements taken at least one week apart on 20 randomly selected models. Following calibration, both operators independently measured all models, repeating the procedure separately.

Digital measurements were performed using a Simhevn Electronic Digital Caliper with a precision of 0.01 mm, measurement range :0-150mm(Fig. 1).Traditional measurement was taken of the cast with the pin point divider and measure on the scales.(Fig 2)

Tooth dimensions were recorded in perpendicular planes, including:

- Maximum mesiodistal widths
- Maximum buccolingual (or bucco-palatal) widths
- Facial axis of the clinical crown
- Arch measurements

These measurements were taken from the first molar on one side to the first molar on the opposite side for both dental arches.(Fig 3)

Arch measurements included:

- Inter-canine distance
- Inter-first premolar distance
- Inter-second premolar distance
- Intermolar distance
- Arch length: measured diagonally from the mesiobuccal cusp tips of the first molars to the mesial contact points of the central incisors
- Arch perimeter : calculated as the sum of two bilateral segments:
    - Segment 1: from the distal surface of the first molar to the mesial contact point of the first premolar[1]
    - Segment 2: from the distal contact of the canine to the mesial contact of the central incisor[1]

For arch measurement average values of above parameters are calculated by operator 1 and operator 2.(Table 1 and Table 2) Arch perimeter was calculated as the sum of bilateral segments from molars to incisors, following guidelines used in earlier morphometric studies.[14]Measurements were statistically analysed using variance formulas, coefficient of reliability, and error variance.[15] For statistical purposes, arch length was analysed as separate left and right segments. Similarly, arch perimeter measurements were divided into anterior/posterior and left/right segments. Repeated-measures ANOVA and three-way MANOVA were used to assess method-based, plane-based, and time-based differences.(Table 3 and Table 4)

Fig 1: Simhevn Electronic Digital Caliper with a precision of 0.01 mm, measurement range :0-150mm

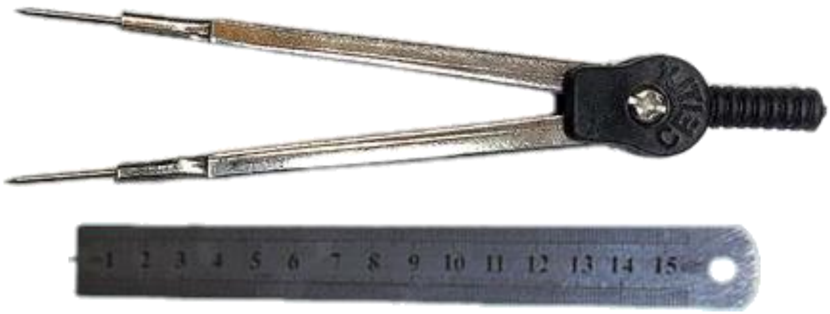

Fig 2: pin point divider and scale

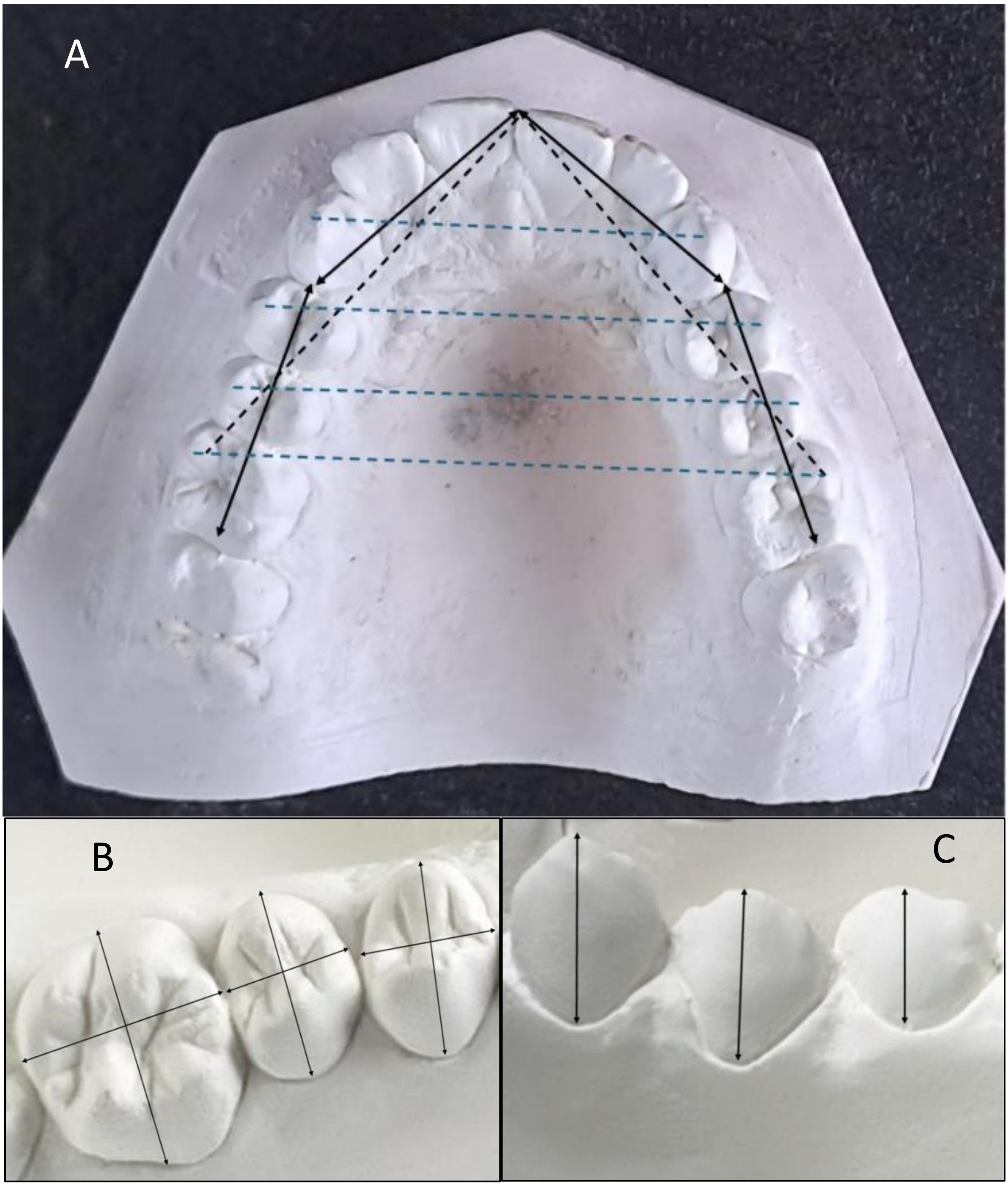

Fig 3:

A. Maxillary plaster model showing linear measurement of intercanine , interfirst premolar, intersecond premolar, intermolar widths(blue dotted lines) ; arch lengths (black dotted lines) ; and arch perimeter segment (solid lines).
B. Closeup view of greatest mesodistal widths and greatest bucco-palatal widths
C. Cloeup view of faxial axis of clinical crown

STATISTICAL ANALYSIS

Accordingly, in this study, linear measurements taken in clinically relevant directions—including arch dimensions, mesiodistal widths, buccolingual widths, and the facial axis of clinical crowns were compared between two methods.

Sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Where S^2 = sample variance
n= number of data points in the sample
xi = each individual data point in the sample
$\bar{x}$ = sample mean (average of the data points)

The **error variance (se^2)** was calculated as half of the variance between the two sets of measurements, providing a reliable estimate of the measurement error for a single observation.[1]

The **total variance (st^2),** which is computed from the series of variance measurements collected across both time points.[1]

The **coefficient of reliability** and **percentage error variance** were determined using the following formulas:[9]

- **Coefficient of reliability** = 1 – (se^2 / st^2) where *se²* is the variance due to random error, and *st²* is the total variance across repeated measurements
- **Percentage error variance** = 100 × (1 – coefficient of reliability)

To assess systematic error (bias), a repeated-measures analysis of variance (ANOVA) was conducted. Repeated observations were treated as within-subject factors, while the measurement method and operator were treated as between-subject factors. A three-way factorial MANOVA was performed with: Independent variables: Method, Plane, Time. Dependent variables: Measurement Mean and Variance. Wilks' Lambda: A measure of how much variance in the dependent variables is *not* explained by the independent variable. Smaller = more significant effect. F-value: Ratio of variance explained to unexplained variance. p-value: If < 0.05, the effect is statistically significant..

Table I - OPERATOR 1

METHOD	PLANE	TIME	MEAN	VARIANCE	ERROR VARIANCE	COEFFICIENT OF RELIABILITY	ERROR VARIANCE(%)
Digital caliper	A	T1	38.875	150.062	0.017	0.999	0.011
		T2	38.809	151.207		1	0
		Difference	0.066	0.034			
	MD	T1	8,100	2.570	0.012	0.995	0.467
		T2	8.001	3.447		1	0
		Difference	0.099	0.025			
	BL	T1	8.200	2.097	0.014	0.993	0.668
		T2	8.162	1.816		1	0
		Difference	0.038	0.028			
	FACC	T1	7.500	2.245	0.020	0.991	0.891
		T2	7.476	2.084		1	0
		Difference	0.024	0.041			
Traditional	A	T1	38.121	150.437	0.056	0.999	0.037

		T2	38.842	150.142		1	0
		Difference	0.729	0.112			
	MD	T1	7.146	2.573	0.015	0.994	0.511
		T2	7.255	3.606		1	0
		Difference	0.109	0.031			
	BL	T1	7.450	1.534	0.014	0.991	0.913
		T2	7.492	1.760		1	0
		Difference	0.042	0.029			
	FACC	T1	6.527	2.236	0.022	0.990	0.984
		T2	6.556	2.419		1	0
		Difference	0.029	0.044			

TABLE II - OPERATOR 2

METHOD	PLANE	TIME	MEAN	VARIANCE	ERROR VARIANCE	COEFFICIENT OF RELIABILITY	ERROR VARIANCE(%)
Digital caliper	A	T1	39.205	148.652	0.020	0.998	0.013
		T2	39.163	149.031		1	0
		Difference	0.042	0.036			
	MD	T1	8.300	2.721	0.011	0.996	0.405
		T2	8.221	3.105		1	0
		Difference	0.079	0.018			
	BL	T1	8.450	1.948	0.012	0.993	0.616
		T2	8.511	2.022		1	0
		Difference	0.061	0.021			
	FACC	T1	7.800	2.163	0.019	0.991	0.879
		T2	7.811	2.275		1	0
		Difference	0.011	0.037			
Traditional	A	T1	38.842	15.101	0.052	0.997	0.034
		T2	38.891	150.879		1	0
		Difference	0.049	0.112			
	MD	T1	7.263	2.684	0.014	0.995	0.521
		T2	7.281	3.221		1	0
		Difference	0.018	0.026			
	BL	T1	7.432	1.663	0.015	0.991	0.902
		T2	7.461	1.712		1	0
		Difference	0.029	0.025			
	FACC	T1	6.568	2.284	0.021	0.990	0.919
		T2	6.581	2.351		1	0
		Difference	0.013	0.041			

RESULT

A total of 20 sets of orthodontic study models met the inclusion criteria. All values were rounded to three decimal places following computation. The normality of the distributions for the values recorded by both operators, using both the digital caliper and traditional method. These analyses confirmed normal distribution for all measured dimensions, including arch lengths, mesiodistal widths, buccolingual widths, and the facial axis of the clinical crown. Intra-operator and inter-operator calibrations demonstrated a high level of consistency, with all arch and tooth size measurements across both methods.

Measurement error variance ranged from 0.012 to 0.020 mm for the digital caliper method and from 0.014 to 0.056 mm for the traditional method (refer to Tables I and II). The calculated reliability coefficients approached 1.0, indicating measurement precision, with error variance accounting for less than 1% of the total variability across all measured dimensions, operators, and time points.

One way ANOVA - There was a statistically significant difference in measurement differences between the two methods ($p < 0.05$). This suggests that the measurement method influenced the consistency in measurements of operator 1 and operator 2. Very large F-statistics indicate very strong evidence of differences between the two methods. (Table III)

MANOVA – (Table IV)
Method ($p < 0.001$) : The measurement method (Digital vs Traditional) significantly affects the combination of Mean and Variance. This means the overall profile of measurements differs significantly between the two methods.

Plane ($p < 0.001$) : The measurement plane (A, MD, BL, FACC) also significantly affects the results, indicating that measurements differ depending on the anatomical plane considered.

Time (T1 vs T2, $p = 0.002$) : There is a significant effect of time, meaning that measurements taken at different times show statistically significant variation.

Method × Plane ($p = 0.003$) : The interaction between method and plane is significant. This suggests that the effects of the measurement method on results varies depending on the plane of measurement.

Method × Time ($p = 0.021$) : The interaction between method and time is significant, indicating that differences between methods also depend on the time point of measurement.

Plane × Time ($p = 0.061$) : This interaction is not statistically significant at the 0.05 level, meaning no strong evidence that effect of plane changes across time points.

Method × Plane × Time ($p = 0.185$) : Not significant. There is no strong combined effect involving all three factors at once.

ONE WAY ANOVA (Table III)

OPERATOR 1

PLANE	F – STATIC	P- VALUE
A	25.7694	0.0012
MD	133.2903	0.0074
BL	628.5536	0.0016
FACC	2528.8984	0.0004

ORERATOR 2

PLANE	F – STATIC	P- VALUE
A	76.3019	0.0013
MD	595.3586	0.0017
BL	937.4450	0.0011
FACC	2090.5310	0.0003

MANOVA (Table IV)

EFFECT	WILKS LAMBDA	F – VALUE	P – VALUE
Method	0.375	9.82	<0.001
Plane	0.095	9.54	<0.001
Time	0.238	7.98	0.002
Method X Plane	0.118	3.92	0.003
Method X Time	0.725	4.51	0.021
Plane X Time	0.441	2.14	0.061
Method X Plane X Time	0.563	1.49	0.185

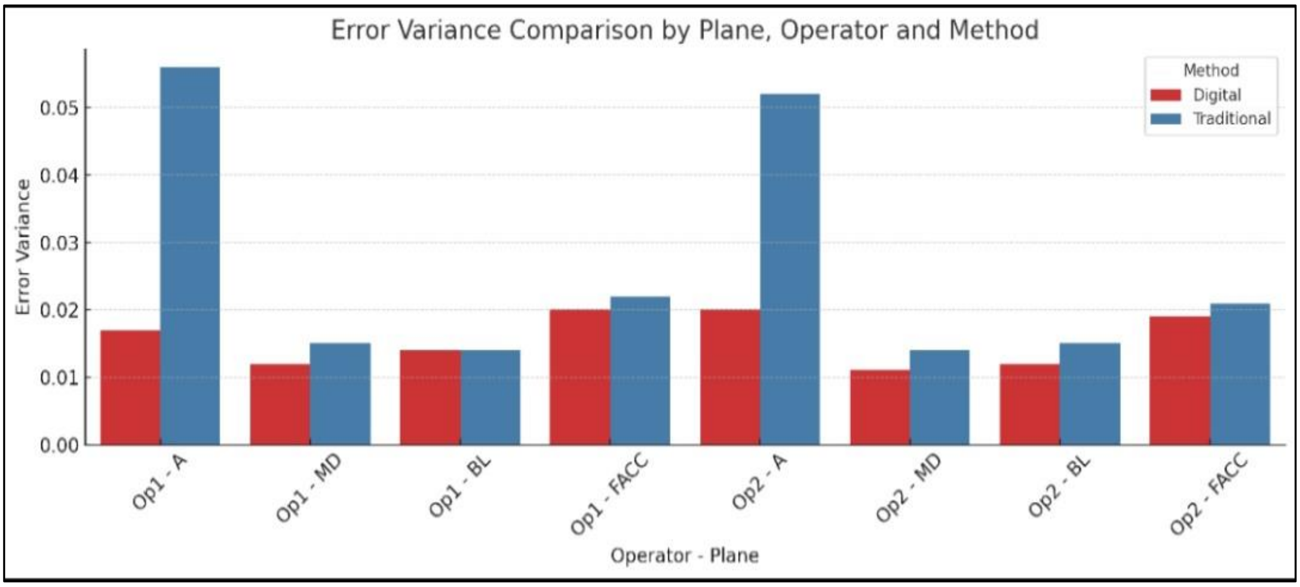

Fig 3 - This chart presents the error variance (a measure of measurement variability) for each plane and operator, again comparing the Digital and Traditional methods.

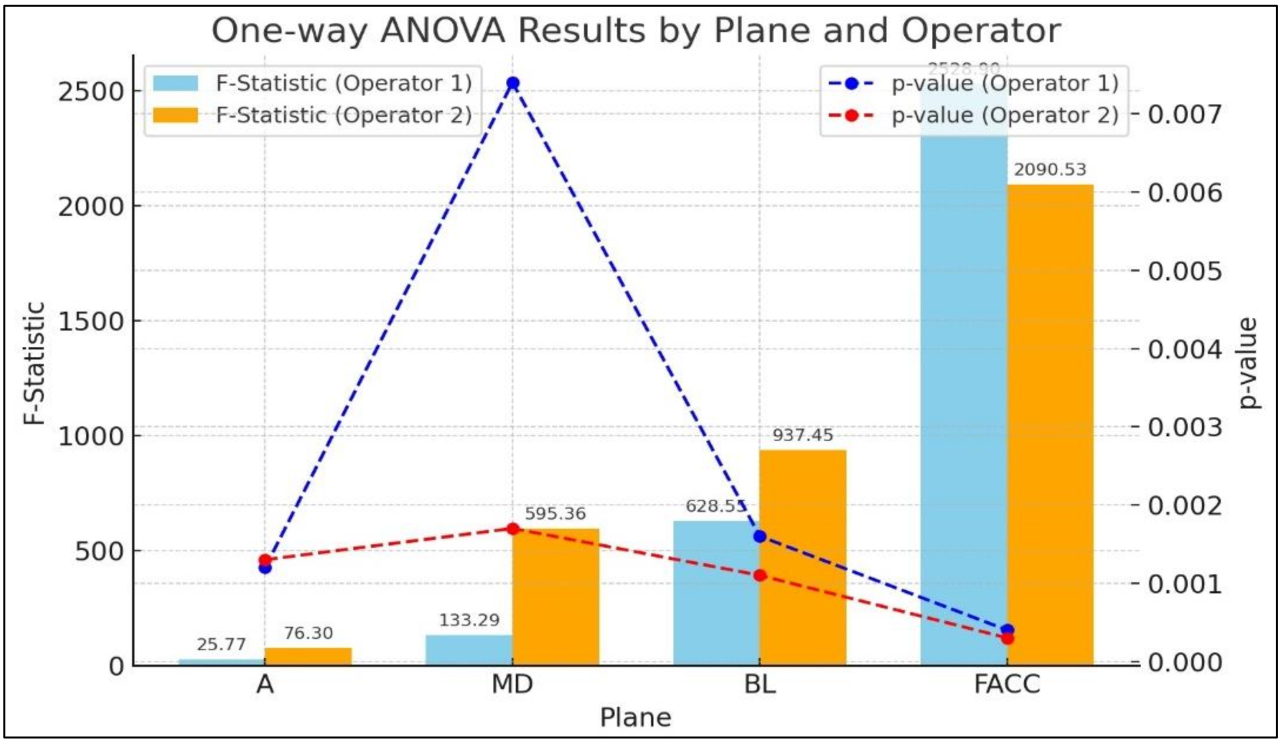

Fig 4 – Combined graph showing Fstatic values and P-values as a line for each plane in ANOVA test

DISCUSSION

Manual tools have been the traditional choice for orthodontic measurements, but their accuracy is often affected by landmark identification and operator pressure.[16] These methods are straight forward and cost-effective but are susceptible to human error, especially in identifying anatomical landmarks and maintaining consistent measurement pressure. Manual methods are generally reliable, their accuracy can be compromised in cases with complex dental anatomies or severe crowding.

Digital calipers provide enhanced precision, often measuring to the nearest 0.01 mm. They reduce subjective errors associated with manual readings. Research indicates that digital calipers yield measurements with high repeatability and minimal variance.

The choice between digital calipers and manual methods should consider factors such as the complexity of the case, available resources, and the need for precision. Digital calipers offer improved accuracy and are less prone to human error, making them suitable for detailed measurements. Manual methods remain valuable, especially in settings where digital tools are unavailable. Ultimately, both methods can provide reliable measurements when used appropriately, and the integration of digital technologies should complement rather than replace traditional techniques.

Although both traditional and digital measuring methods can have errors, digital calipers are currently regarded as the benchmark for comparing measurements between plaster and digital study models. When transitioning to a newer system, it is important to ensure that the chosen tools are repeatability, and provide flexible viewing options for assessing casts individually as well as in occlusion.

Measurement errors can be random or systematic. Training and calibration can reduce but not eliminate human errors.[17]Digital calipers minimize these errors and are considered a practical standard, especially for comparing physical and virtual models.[18]Systematic error is the bias that may result from lack of calibration or magnification of the system, causing it to consistently give higher or lower values.[1] Although extensive training and calibration were carried out, fatigue from measuring samples and lack of experience may still cause errors.[1] Errors are inevitable because the measurements were made by human operators, and they may be improved by experience.[1]

This study used multiple statistical tools, including reliability coefficients, repeated-measures ANOVA, and MANOVA to evaluate agreement and error variance. Error variance remained below 1%, well within clinically acceptable limits[19] Random errors contribute to an increase in the total measurement variance (st^2), and the coefficient of reliability as a way to quantify the impact of these errors on the consistency between repeated measurements. A coefficient of reliability value of 1 indicates the absence of random error, reflecting perfect reliability and accurate representation of the measurement target.[1] Reliability coefficients very close to 1, indicating negligible error variance regardless of whether the total variance from the first or second set of measurements was considered.[1]

While the acceptable threshold for error variance has been suggested to be less than 10%, the results of this study, with error variances below 1%, suggest that random errors were minimal and the precision of the measurements was excellent.

To assess measurement reliability and consistency between the two methods, a Repeated Measures ANOVA and MANOVA (Multivariate Analysis of Variance) were applied. Repeated Measures ANOVA was used to detect significant differences within each method across repeated trials (T1 and T2). MANOVA was employed to evaluate the effect of measurement method, anatomical plane (A, MD, BL, FACC), and time point on multiple dependent variables: Mean and Variance.

Advancements in 3D scanning and modelling have introduced non-contact measurement techniques. These methods allow for comprehensive analysis of dental arches and are particularly useful in complex cases. However, their accuracy is contingent on the quality of the scan and the software's ability to accurately render dental structures. Some studies have reported minor discrepancies between digital and manual measurements, but these differences are often clinically insignificant.[20]

CONCLUSION

This study systematically compared the measurements of digital calipers versus traditional manual methods in measuring orthodontic study models. The results demonstrated that both methods are capable of delivering accurate and consistent measurements, with digital calipers showing slightly superior precision and repeatability. Statistical analyses, including ANOVA and MANOVA, confirmed significant method-based differences.

The digital caliper consistently exhibited lower error variance and higher reliability coefficients across operators and time points, underscoring its efficiency in minimizing operator-dependent variability. Nonetheless, the manual method remains valuable due to its simplicity, cost-effectiveness, and clinical familiarity, especially in resource-limited settings.Ultimately, both tools can be reliably employed for orthodontic measurements when used with appropriate training and calibration. The choice of method should be guided by clinical context, operator expertise, and the precision requirements of each case. As digital technologies continue to advance, their integration should complement, rather than replace, conventional techniques to enhance the quality and accuracy of orthodontic diagnostics and treatment planning.

REFERENCES

1. Assessing agreement in measurement of orthodontic study models. Digital caliper on plaster models vs 3-dimensional software on models scanned by structured light scanner. Wan Hassan WN, Othman SA, Chan CS, Ahmad R, AliSN, Abd Rohim A. Am J Orthod Dentofacial Orthop 2016 Nov: 150(5) : 886-895.

2. Houston WJB. The analysis of errors in orthodontic measurements. Am J Orthod. 1983;83(5):382–90.
3. Zilberman O, Huggare J, Parikakis KA. Evaluation of the validity of tooth size and arch width measurements using conventional and three-dimensional virtual orthodontic models. Angle Orthod. 2003;73(3):301–6.
4. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. Arch Orofac Sci. 2017;12(1):1–11.
5. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 6th ed. Boston: Pearson Education; 2013.
6. Bootvong K, Liu Z, McGrath C, Hagg U, Wong RWK, Bendeus M, et al. Virtual model analysis as an alternative approach to plaster model analysis: Reliability and validity. Eur J Orthod. 2010;32(5):589–95.
7. Rossini G, Parrini S, Castroflorio T, Deregibus A, Debernardi CL. Diagnostic accuracy and measurement sensitivity of digital models for orthodontic purposes: A systematic review. Am J Orthod Dentofacial Orthop. 2016;149(2):161–70.
8. Santoro M, Galkin S, Teredesai M, Nicolay OF, Cangialosi TJ. Comparison of measurements made on digital and plaster models. Am J Orthod Dentofacial Orthop. 2003;124(1):101–5.
9. Lombardo L, Sgarbanti C, D’Este G, Barbato E. Accuracy of digital and traditional orthodontic measurements. Prog Orthod. 2020;21(1):20.
10. Zaganjori H, Maal TJJ, Bronkhorst EM, Langenbach GEJ, Kuijpers-Jagtman AM, Schols JGJH. Accuracy of linear measurements on 3D virtual models of human skulls: A systematic review. PLoS One. 2018;13(12):e0208880.
11. Naidu D, Freer TJ. Validity, reliability, and reproducibility of the iOC intraoral scanner: A comparison of tooth widths and Bolton ratios. Am J Orthod Dentofacial Orthop. 2013;144(2):304–10.
12. Ayoub AF, Wray D, Owen CP, Moos KF. A comparison of the accuracy of measurements performed on plaster models and 3D images for planning orthognathic surgery. Int J Oral Maxillofac Surg. 1998;27(2):138–43.
13. Chiu AB, Clark JR, Halazonetis DJ. Comparison of hand-traced and computer-aided cephalometric analysis. Am J Orthod Dentofacial Orthop. 2015;148(5):820–4.
14. Al-Khateeb SN, Alhaija ESJ. Tooth size discrepancies and arch parameters among different malocclusions in a Jordanian sample. Angle Orthod. 2006;76(3):459–65.
15. Dahlberg G. Statistical Methods for Medical and Biological Students. London: George Allen & Unwin Ltd; 1940.
16. Rangel FA, Maurette PE, Allgayer S, Lagravère MO. Reproducibility of linear measurements obtained from cone-beam computed tomography images of human dry skulls. J Craniofac Surg. 2010;21(5):1503–7.
17. Mullen SR, Martin CA, Ngan P, Gladwin M. Accuracy of space analysis with emodels and plaster models. Am J Orthod Dentofacial Orthop. 2007;132(3):346–52.
18. Houston WJB. Accuracy of measurement of orthodontic study models. Br J Orthod. 1981;8(3):149–52.
19. Luu NS, Mandich MA, Heo G, Flores-Mir C. Evidence supporting the use of digital orthodontic models: A systematic review. Angle Orthod. 2012;82(4):629–34.
20. Stevens DR, Flores-Mir C, Nebbe B, Heo G, Major PW. Validity, reliability, and reproducibility of plaster vs digital study models: Comparison of peer assessment rating and Bolton analysis. Am J Orthod Dentofacial Orthop. 2006;129(6):794–803.