BMC Oral Health

RESEARCH Open Access

# Reproducibility and reliability of dental arch measurements: comparing of manual, digital, and app-based methods

Alfonso Alvarado-Lorenzo[1], Laura Antonio-Zancajo[1*], Adrián Curto[1], Daniele Garcovich[2] and Laura Criado-Pérez[1]

## Abstract

**Background** Knowing the available dental space in a patient is crucial for orthodontists to develop a good treatment plan.

**Objectives** To compare and evaluate the reliability, reproducibility, and accuracy of three measurement methods on models: conventional, mobile app, and digital software. Materials and methods: Maxillary and mandibular dental plaster models of 20 subjects with permanent dentition were analyzed. Three measurements were taken: manually using a digital caliper, using Nemocast 3D® software, and using the mobile app "Ruler." These were performed by the same observer at the same initial time (T1) and repeated after 15 days (T2).

**Results** The three methods measured each of the quadrants and each of the dental pieces similarly, with no significant differences found. Significant differences were found between the different measurement times in the manual and app ruler measurement groups. Intra-observer consistency was also observed through Pearson correlations. The measurement from the mobile app "Ruler" presented lower values compared to the other two models, especially with the measurements taken for each quadrant.

**Conclusion** The three methods used for measuring dentoskeletal discrepancy are similar in terms of results. However, Nemocast 3D software showed greater reproducibility and reliability compared to the other two methods presenting consistency in the results measured at both time points, which may be useful for our diagnoses.

**Trial registration** The study was conducted according to the ethical standards set by the Declaration of Helsinki for Biomedical Research and the protocol was approved by the Bioethics Committee of the University of Salamanca with number 23/1074.

**Keywords** Plaster model, Space analysis, Digital study models, Smartphone dental apps, Digital orthodontic measurement, Orthodontic artificial intelligence

*Correspondence:
Laura Antonio-Zancajo
lantoniozancajo@yahoo.es
[1]Department of Oral Surgery, Universidad de Salamanca, Salamanca 37007, Spain
[2]Department of Dentistry, Universidad Europea de Valencia, Valencia 46010, Spain

## Introduction

Orthodontic study models, along with the patient's history, radiological examination, and clinical examination, allow the professional to achieve an orthodontic diagnosis and accurate treatment planning [1–3]. Dental arch discrepancy ) results from the difference between the arc length, called "available space," and the mesiodistal length of each tooth mesial to the first molar, called "required space." It is diagnosed by observing a series of patient symptoms, clinical signs, and radiological signs. This discrepancy between tooth size and jaw dimension is the most common reason patients visit orthodontic consultations. Therefore, more digitalized and computerized tools are increasingly being implemented to facilitate its evaluation and thus allow for better orthodontic treatment planning [4–6].

Traditionally, orthodontists have used plaster dental models and calipers (manual or digital) to perform tooth width measurements. With adequate impression and casting techniques, these models can provide precise reproducibility of dentition and surrounding structures [7–10]. However, these models have certain limitations, such as the risk of chipping, abrasion, or breakage, creating the need for storage space and associated costs [7, 8, 10].

Currently, various methods for obtaining digital models and/or measurements have emerged, as they have several advantages such as easy access and storage, in addition to facilitating communication between different professionals [11–13]. After the introduction of these new techniques in the market, multiple studies have been conducted comparing linear measurements of digitalized models versus manual measurement in plaster models [7, 13–18].

The advantages of these digital methods largely rely on the reliability of both the software and the hardware of the digitization, but it is also necessary for the operator to master the measurement technique [19, 20]. There is an increasing availability of orthodontic applications identified in Apple App Store and Google Play Store [21, 22]. Among these specific orthodontic tools are some related to model analysis, which function through photogrammetry, a process in which overlaid images produce 2D or 3D digital models [23, 24]. Other authors have implemented the use of cbct or intraoral scanner for the manufacture of 3d models [20, 25]. On the other hand, various research studies have concluded that mobile health applications, through the use of wireless devices and smartphones, are spreading very quickly, contributing to improved patient outcomes, advancement in health research, and healthcare services [26, 27].

Despite this current trend of using digital tools and devices for dental measurement evaluation, there are no studies that thoroughly analyze and compare the accuracy of orthodontic measurements (such as the discrepancy in arch length, using mobile applications and dental model analysis software). We have not found in the literature any studies that analyze these three methods with the same purpose and very few analyze applications like the app ruler. The aim of this study is to compare the reproducibility, reliability, and accuracy of different methods of analyzing dental arch discrepancy using Nemocast software, the mobile app "Ruler," and the traditional manual method. The null hypothesis of this work is that there are no differences in terms of reliability and reproducibility among the three methods of orthodontic model measurement.

## Material and method

### Sample size calculation and participants

For this study, a total of 20 participants were used, 11 males (55%) and 9 females (45%). This sample size is similar to that of other analyzed articles 4,7,17,29]. The average age of the patients was 14.5 years, with a range between 13 and 18 years. The sample size required for this study was estimated using GPower v. 3.1.9.6, assuming a significance level of $\alpha=0.05$, a statistical power of 95%, and an effect size $d=0.80$, resulting in 19 participants.

The inclusion criteria were as follows: adolescent patients, teeth with normal dental morphology, permanent dentition, teeth without wear and abrasion. The exclusion criteria were: patients with prostheses, teeth with restorations, caries, morphological alterations, or those in which interproximal reduction (IPR) had been performed previously.

The study was conducted according to the ethical standards set by the Declaration of Helsinki for Biomedical Research and the protocol was approved by the Bioethics Committee of the University. All the patients who participated in the study were informed about the procedures to be followed and were provided with informed consent for their approval.

### Study design

Initially, alginate impressions were taken from each of the patients (Cavex Impressional; Cavex Holland BV, Haarlem, The Netherlands) and immediately after, the plaster model was obtained with white stone (Ryoka Dental; Mie-Ken, Japan) with proportions indicated by the manufacturer to ensure as much as possible that variations during setting due to misuse of the proportions are minimized. For each patient in the sample, three sets of measurements were taken at two different time points (T1 and T2). T1 was the initial measurement moment, and T2 was 15 days after the previous measurement. All measurements were performed by the same operator (see Fig. 1), taking as a reference the mesial faces of the first

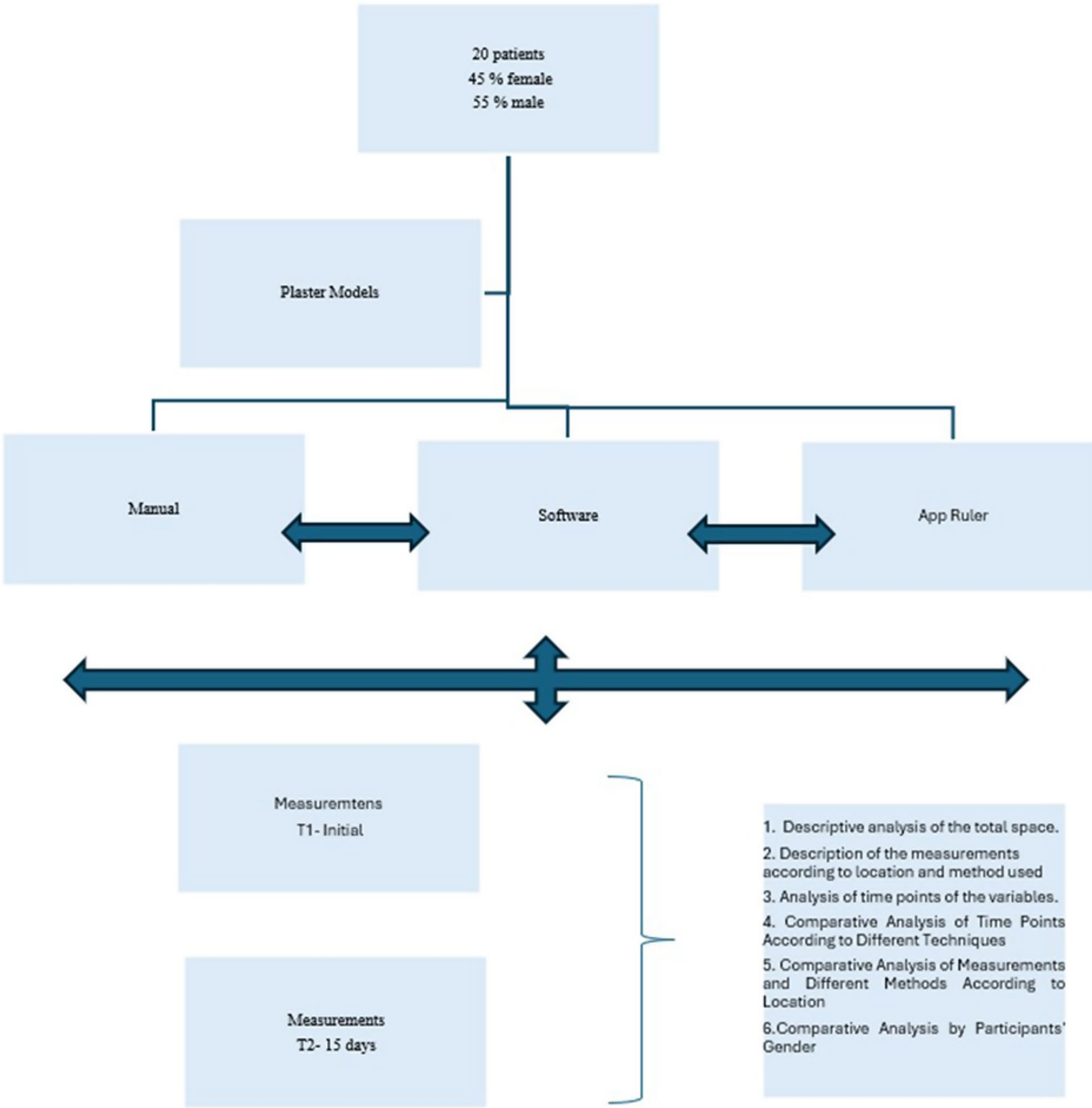


**Fig. 1** Illustration of the flowchart of measurement with different techniques

molars. Each measurement was recorded using a different methodology: manual method with measuring caliper, digital method through Nemocast 3D software, and digital method through the mobile app "Ruler".

### *Manual method*

The manual caliper is a measuring instrument designed to measure and check small distances 4. For the manual measurement of the models used in this study, a Mitutoyo vernier digital caliper (Mitutoyo Co, Kanagawa, Japan) with modified tips and 0.01 mm precision was used (Fig. 2a and b).

The digital caliper displays the measurement on a liquid crystal display mounted on the sliding part and employs the principle of the digital potentiometer. Measurements were conducted with a lateral approach, that is, viewing from the occlusal perspective of the study model, placing the tips of the caliper at a right angle to the axial axis of the tooth being measured and with the handle of the instrument parallel to the occlusal surface of the same.

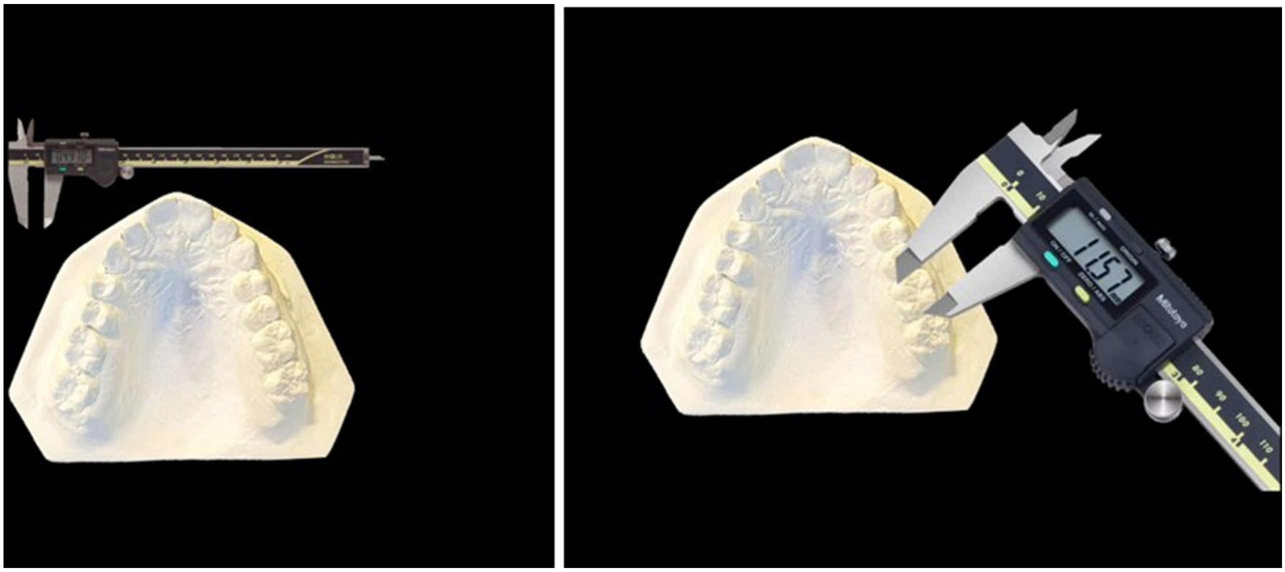

**Fig. 2** (**a**, **b**) Mitutoyo vernier digital caliper

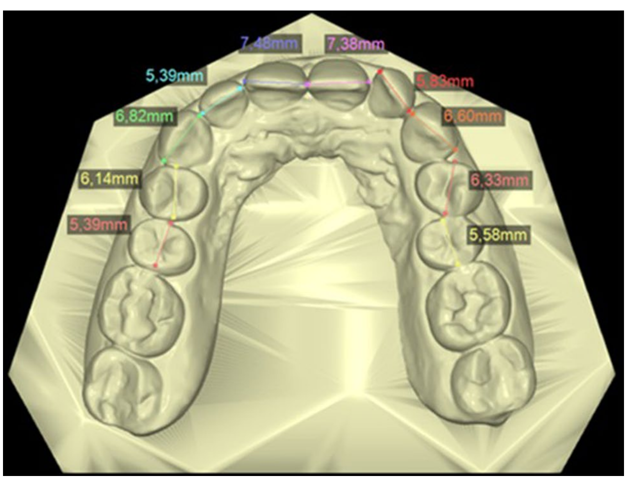


**Fig. 3** Measurement performed using Nemocast 3D

### *Digital method using orthodontic diagnostic software*

In this case, Nemocast 3D software version 11.3.1, 2013 (Nemotec, Madrid, Spain) was used. First, the models were digitized and converted to STL format. Data were entered and digital models were prepared with Nemocast 3D software (Fig. 3), following an ordered sequence of steps:

- Orientation in the three spatial axes X, Y, Z.
- Placement of occlusal planes and midline.
- Marking of the mesio-distal points of each of the teeth included in the analysis.
- The models were based according to ABO standards (American Board of Orthodontics).
- After obtaining the digitized model, the software automatically provided any data required for diagnosis.

### *Digital method using the mobile app "Ruler"*

Regarding measurement via mobile phone, the "Ruler" app was used (Nguyen Minh, Vietnam). This app included documents on how to operate the program, capture images, and perform the measurement, and which were properly consulted for correct use. For calibration of the app, an external physical ruler and a credit card were used, which were compared with the virtual rulers and credit cards of the app. After that, a dimension pointer displayed the measured values of the object in question, from which the required space and available space were obtained (Fig. 4a and b).

### Statistical analysis

Finally, statistical analyzes of the data were performed using IBM SPSS Statistics software (version 29.0, IBM SPSS Inc., Chicago, IL, USA). The main descriptives (mean and standard deviation) were obtained for each method, quadrant and time point. The reliability of the methods was analyzed using Pearson correlations of the total scores at T1 and T2. The Pearson coefficient was used for the quantitative variables. To compare the measurements of each method separately at T1 and T2, several Student's t-tests for paired samples were calculated for each tooth, quadrant, and arch. Values of the p value below 0.01 ($p<0.01$) were considered to be highly significant results and p values below 0.05 ($p<0.05$) were considered significant. Regarding the comparison between men and women, various Student's t tests were calculated for independent samples for each tooth, quadrant and arch.

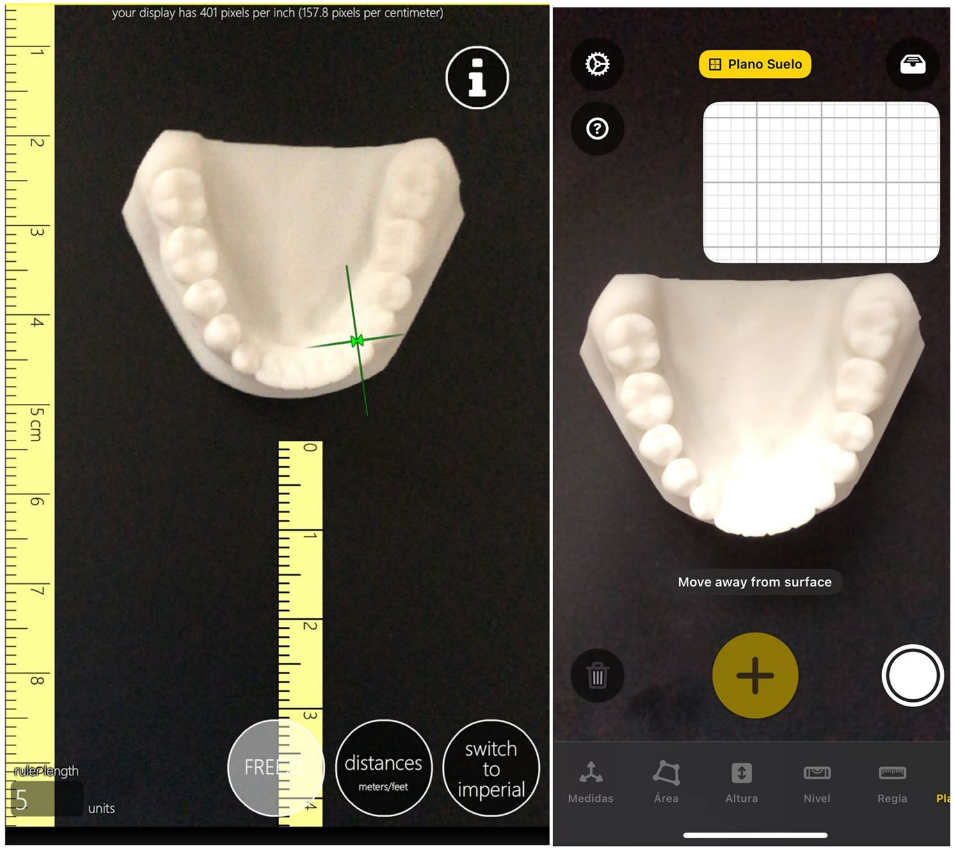


**Fig. 4** (**a**, **b**) Measurement of models by ruler application

**Table 1** Mean (Standard Deviation) of total measured space for each method and time point

| Method | Time Point | | | | | |
|---|---|---|---|---|---|---|
| | $T_1$ | | | $T_2$ | | |
| | **Avail. space** | **Req. space** | **Dif. (mm)** | **Avail. space** | **Req. space** | **Dif. (mm)** |
| Manual | 71.59 (6.91) | 72.79 (6.33) | -1.20 (3.63) | 71.92 (7.04) | 72.41 (6.02) | -0.49 (3.58) |
| Software | 68.48 (6.03) | 68.85 (5.33) | -0.37 (3.19) | 68.69 (7.04) | 68.80 (5.11) | -0.11 (3.28) |
| App Ruler | 69.93 (7.04) | 71.45 (6.46) | -1.52 (4.66) | 70.85 (6.47) | 71.18 (6.14) | -0.33 (4) |

Avail. space = Available space. Req. space = Required space. Dif. = Difference between available space and required space measured in millimeters (mm)

## Results

### Descriptive analysis of the total space measured according to the method used

After analyzing the total available and necessary space in the different models at the first time point (T1), we observed that the three methods used did not coincide in obtaining the same measurement of dental arch discrepancy. The same occurs if we analyze the results after the 15 days of the first measurement (T2), however, these differences are smaller among the three methods (Table 1).

In the two time points analyzed, we observed that the measurements obtained in terms of dental arch discrepancy were lower using the software (Table 1).

### Descriptive analysis of the measurements obtained according to location and method used

Analyzing the data obtained according to location in the dental arch, we found that negative dental arch discrepancy is greater in anterior locations on both the left and right sides, especially using the “Ruler” app at time point T1 (-0.68 mm on the right anterior and −0.5 mm on the left anterior). At time point T2, the greatest negative discrepancy is observed in the right anterior location through manual measurement (-0.45 mm) (Table 2).

Measurements of dental arch discrepancy+were obtained in the left posterior location both in measurement with software (+0.39 mm) and Ruler app (+0.09 mm) at time point T2 (Table 2).

**Table 2** Mean (standard deviation) of space measured for each method, quadrant, and time point

| Sectors | Method | Time Point | | | | | |
|---|---|---|---|---|---|---|---|
| | | $T_1$ | | | $T_2$ | | |
| | | Avail. space | Req. space | Dif (mm) | Avail. space | Req. space | Dif. (mm) |
| Posterior Right | Manual | 21.88 (1.67) | 22.05 (1.20) | -0.17 (1.39) | 22.01 (1.56) | 21.84 (1.20) | 0.17 (1.09) |
| | Software | 21.26 (1.35) | 21.39 (0.99) | -0.13 (1.15) | 21.21 (1.29) | 21.25 (0.94) | -0.04 (1.12) |
| | App Ruler | 21.16 (1.47) | 21.42 (1.24) | -0.26 (1.51) | 21.54 (2.04) | 21.56 (1.35) | -0.02 (1.75) |
| Anterior Right | Manual | 14.10 (2.65) | 14.36 (2.62) | -0.26 (1.10) | 14.14 (2.72) | 14.35 (2.60) | -0.21 (1.15) |
| | Software | 13.02 (2.76) | 13.32 (2.16) | -0.30 (1.36) | 13.12 (2.55) | 13.39 (2.20) | -0.27 (1.15) |
| | App Ruler | 13.59 (2.65) | 14.27 (2.38) | -0.68 (0.44) | 13.89 (2.58) | 14.02 (2.22) | -0.13 (1.77) |
| Anterior Left | Manual | 13.79 (2.67) | 14.21 (2.40) | -0.42 (1.50) | 13.90 (2.70) | 14.35 (2.25) | -0.45 (1.58) |
| | Software | 12.82 | 13.11 (2.29) | -0.29 (1.45) | 12.90 (2.69) | 13.09 (2.16) | -0.19 (1.49) |
| | App Ruler | 13.67 (2.78) | 14.17 (2.56) | -0.50 (1.67) | 13.78 (2.56) | 14.04 (2.33) | -0.27 (1.92) |
| Posterior Left | Manual | 21.81 (1.48) | 22.07 (1.22) | -0.26 (1.15) | 21.84 (1.53) | 21.88 (1.21) | -0.04 (1.26) |
| | Software | 21.38 (1) | 21.03 (1.10) | 0.35 (0.68) | 21.46 (1.03) | 21.07 (1) | +0.39 (0.75) |
| | App Ruler | 21.50 (1.66) | 21.58 (1.48) | -0.08 (1.36) | 21.64 (1.97) | 21.55 (1.55) | +0.09 (1.56) |

Avail. space = Available space. Req. space = Required space. Dif. = Difference between available space and required space measured in millimeters (mm)

**Table 3** Comparison between measurements at the two temporal moments, based on the measurement method employed

| Method | $M_{T1}$ | $M_{T2}$ | d | *p* |
|---|---|---|---|---|
| Manual | -1.206 | -0.497 | 0.709 | < 0.001** |
| Software | -0.367 | -0.307 | 0.259 | 0.137 |
| App Ruler | -1.522 | -0.333 | 1.19 | < 0.001** |

MT1; MT2 = Means at T1 and T2, d = effect size, mean difference (T2-T1), p: p-value

** Statistically significant results ($p < 0.01$)

**Table 4** Correlations between the scores of each model at each time point

| Method | Correlation (Sector scores) | Correlation (Total score) |
|---|---|---|
| Manual | 0.868** | 0.919** |
| Software | 0.823** | 0.879** |
| App Ruler | 0.414** | 0.747** |

** $p < 0.01$

## Analysis of Time points of the variables

No significant differences were found when comparing the means obtained by the three methods in each quadrant, showing the consistency of the measurement of the three methods. Regarding the effect of the time of measurement, only the "Nemocast" software showed consistency in the average measurement over time ($p = 0.137$), with significant differences found between T1 and T2 in the manual method ($p < 0.001$) and the "Ruler" app ($p < 0.001$) (Table 3).

Additionally, Pearson correlations between the measurements at T1 and T2 (scores by quadrant and total) were analyzed, obtaining significant differences (Table 4). The measurement from the mobile app "Ruler" showed lower values than the other two models, especially with the measurements taken for each quadrant, however, the values for the manual method and the Nemocast software were acceptable, which gives us an idea of the reproducibility.

## Comparative analysis of time points according to different techniques

Table 5 presents the comparisons between pairs of methods for the total scores at each time point. In the first measurement (T1), significant differences were found between the measurement performed by the "Nemocast" software and the "Ruler" app (d = 1.15; $p = 0.027$; 95% CI [0.10–2.20]). In the second measurement (T2), no differences were found between the methods used. Subsequent analyses were repeated using the scores by quadrant as the unit of measure, finding no differences based on the method neither at T1 nor at T2, Therefore, the manual method and the Nemocast software have the best results between them.

**Table 5** Levels of significance of mean differences between models at each time point

| Comparision | Time Point | | | | | |
|---|---|---|---|---|---|---|
| | $T_1$ | | | $T_2$ | | |
| | d | p | IC95% | d | p | IC95% |
| Manual-Software | -0.84 | 0.073 | -1.74–0.06 | -0.39 | 0.603 | -1.34–0.57 |
| Manual-App Ruler | 0.32 | 0.777 | -0.78–1.42 | -0.16 | 0.913 | -1.12–0.79 |
| Software-App Ruler | 1.15 | 0.027 | 0.10–2.20 | 0.22 | 0.845 | 0.73–1.18 |

**Table 6** Comparison of the difference in means in T1 and T2 in each tooth and upper arch, for each measurement method used

| Tooth | Manual | | | Software | | | App Ruler | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Media** | **DT** | ***p*** | **Media** | **DT** | ***p*** | **Media** | **DT** | ***p*** |
| 15 | -0,125 | 0,315 | 0,093 | -0,082 | 0,314 | 0,260 | -0,025 | 0,415 | 0,791 |
| 14 | -0,030 | 0,330 | 0,689 | -0,012 | 0,256 | 0,843 | -0,090 | 0,440 | 0,372 |
| 13 | -0,071 | 0,373 | 0,406 | -0,056 | 0,353 | 0,486 | 0,155 | 0,493 | 0,175 |
| 12 | 0,065 | 0,446 | 0,526 | 0,071 | 0,414 | 0,453 | -0,070 | 0,511 | 0,547 |
| 11 | -0,089 | 0,261 | 0,146 | 0,014 | 0,274 | 0,828 | -0,295 | 0,666 | 0,025* |
| 21 | -0,049 | 0,376 | 0,571 | -0,056 | 0,255 | 0,342 | -0,270 | 0,790 | 0,143 |
| 22 | 0,059 | 0,290 | 0,379 | -0,099 | 0,270 | 0,118 | -0,086 | 0,584 | 0,523 |
| 23 | -0,107 | 0,337 | 0,172 | -0,067 | 0,637 | 0,644 | 0,005 | 0,749 | 0,976 |
| 24 | -0,081 | 0,302 | 0,245 | 0,029 | 0,236 | 0,589 | -0,170 | 0,606 | 0,225 |
| 25 | -0,025 | 0,289 | 0,703 | -0,004 | 0,346 | 0,954 | 0,195 | 0,454 | 0,070 |
| **Maxilar** | | | | | | | | | |
| AAL | -0,014 | 0,797 | 0,912 | -0,070 | 0,813 | 0,589 | -0,720 | 2,144 | 0,040* |
| IAL | -0,225 | 0,580 | 0,004** | -0,149 | 0,438 | 0,011* | 0,040 | 1,081 | 0,775 |
| DAL | -0,213 | 0,485 | 0,001*** | -0,043 | 0,915 | 0,720 | -0,010 | 1,565 | 0,961 |

Related-samples t-test (*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$). AAL: Anterior Arch Length; IAL: Left Arch Length; DAL: Right Arch Length

**Table 7** Comparison of the difference in means in T1 and T2 in each tooth and lower arch for each measurement method used

| Tooth | Manual | | | Software | | | App Ruler | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Media** | **DT** | ***p*** | **Media** | **DT** | ***p*** | **Media** | **DT** | ***p*** |
| 35 | 0,005 | 0,344 | 0,954 | 0,445 | 0,268 | 0,467 | 0,040 | 0,417 | 0,673 |
| 34 | -0,077 | 0,244 | 0,174 | -0,059 | 0,306 | 0,399 | 0,055 | 0,330 | 0,465 |
| 33 | -0,137 | 0,321 | 0,073 | -0,108 | 0,328 | 0,158 | 0,135 | 0,456 | 0,201 |
| 32 | 0,003 | 0,200 | 0,956 | 0,072 | 0,363 | 0,390 | -0,04 | 0,490 | 0,719 |
| 31 | 0,006 | 0,196 | 0,892 | -0,012 | 0,285 | 0,853 | -0,045 | 0,450 | 0,660 |
| 41 | 0,132 | 0,315 | 0,077 | 0,105 | 0,296 | 0,129 | -0,045 | 0,679 | 0,770 |
| 42 | 0,138 | 0,426 | 0,164 | 0,008 | 0,267 | 0,901 | 0,145 | 0,695 | 0,363 |
| 43 | -0,039 | 0,232 | 0,468 | 0,018 | 0,436 | 0,855 | -0,020 | 0,588 | 0,881 |
| 44 | -0,245 | 0,223 | 0,629 | 0,063 | 0,280 | 0,326 | 0,025 | 0,459 | 0,810 |
| 45 | -0,111 | 0,326 | 0,144 | 0,029 | 0,352 | 0,717 | -0,040 | 0,438 | 0,688 |
| **Jaw** | | | | | | | | | |
| AAL | 0,278 | 0,740 | 0,022* | 0,172 | 0,742 | 0,151 | -0,035 | 1,561 | 0,888 |
| IAL | -0,209 | 0,606 | 0,010** | -0,123 | 0,597 | 0,117 | 0,230 | 0,782 | 0,026* |
| DAL | -0,174 | 0,527 | 0,013* | 0,110 | 0,547 | 0,193 | -0,040 | 1,261 | 0,807 |

Related-samples t-test (*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$). AAL: Anterior arch length; IAL: Left arch length; DAL: Right arch length

### Comparative analysis of measurements and different methods according to location

Analyzing the measurements according to the dental piece, it is observed that there are no statistically significant differences between the measurement methods, except in one case (piece 1.1 in "App Ruler") (Table 6).

By sectors, we can distinguish between the upper and lower arches. In the upper arch, the manual method presents statistically significant differences in the measurements of the left ($p \leq 0.01$) and right ($p \leq 0.001$) sectors, showing greater validity in the measurement both by software and by mobile app (Table 6).

In the case of the lower arch, again significant differences appear in all sectors when measured with the manual method, with the "Nemocast" software showing measurements with greater temporal reliability (Table 7).

Finally, analyzing the measurements in each arch (Table 8), it is observed that the measurement by the "Nemocast" software is more reliable in the upper arch (Table 8), while the "Ruler" app and the manual method present more reliable scores in the lower arch (Table 9).

## Discussion

Advanced computerized systems, digital models, and mobile applications offer promising clinical solutions for orthodontists. However, before deciding to introduce a new method into daily practice, it is necessary to ensure that this new measurement method is easily accessible, accurate, repeatable, and reproducible, just like methods currently considered more traditional. Therefore, the purpose of this study has been to compare the reliability and validity of orthodontic measurements using

**Table 8** Comparison of the difference in means at T1 and T2 in the upper arch

| | Manual | | | Software | | | App Ruler | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Media** | **DT** | ***p*** | **Media** | **DT** | ***p*** | **Media** | **DT** | ***p*** |
| ED | 0.137 | 0.446 | 0.000*** | 0.029 | 0.410 | 0.319 | 0.176 | 1.654 | 0.134 |
| EN | -0.134 | 0.527 | 0.000*** | -0.071 | 0.628 | 0.109 | -0.135 | 1.264 | 0.133 |
| DIF | 0.271 | 0.682 | 0.000*** | 0.100 | 0.765 | 0.065 | 0.311 | 1.217 | 0.000*** |

Paired t-test (*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$). TD: Total available space (sum of each sector); TN: Total necessary space (sum of each sector); DIF: Difference between total available space and total necessary space

**Table 9** Comparison of the mean difference at T1 and T2 in the lower arch

| | Manual | | | Software | | | App Ruler | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Media** | **DT** | ***p*** | **Media** | **DT** | ***p*** | **Media** | **DT** | ***p*** |
| ED | 0.016 | 0.482 | 0.633 | 0.060 | 0.378 | 0.026* | 0.297 | 2.295 | 0.069 |
| EN | -0.059 | 0.572 | 0.144 | 0.031 | 0.581 | 0.457 | 0.050 | 1.029 | 0.493 |
| DIF | 0.076 | 0.664 | 0.109 | 0.029 | 0.672 | 0.541 | 0.247 | 2.222 | 0.118 |

Paired t-test (*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$). TD: Total available space (sum of each sector); TN: Total necessary space (sum of each sector); DIFF: Difference between total available space and total necessary space

"Nemocast" software, the "Ruler" mobile app, and the conventional method using a digital caliper.

Although manual measurements of the study model are often considered the standard, the accuracy of tooth width values measured with different methods depends both on the accuracy of the dentition replication process and on the consistency and accuracy of the measurement process. In theory, when using orthodontic plaster models, the expansion caused by the setting of dental plaster is compensated by the contraction that occurs in the impression material [29]. Consequently, the three most important factors that could significantly alter the accuracy of measurements made using calipers on physical study models are: the challenging access in dentitions with crowding, difficulties in locating distal and mesial reference points due to aberrant interproximal dental anatomy, and inconsistencies in carrying out established measurement protocols [26, 29].

Regarding the standard deviations of the total space measured for each method and time point, the values obtained by the mobile app "Ruler" (-1.52 mm at T1 and −0.33 mm at T2) were the closest to the manual method (-1.20 mm at T1 and −0.49 mm at T2). The values obtained by the software were the lowest at both time points; a finding consistent with the studies by Grünheid T et al. [7] and Nalcaci et al. [17], where similar results were obtained, showing that the values obtained by software, in digital models, were always lower. However, in the study conducted by Koretsi V et al. [19], greater mesiodistal sizes were obtained in digital measurements compared to manual ones.

Regarding the measurements of each quadrant, no statistically significant differences were found. No statistically significant differences were obtained in individual measurements at the level of each dental piece except for piece 11 with the "Ruler" app method. Therefore, it could be concluded that the three methods measured each of the quadrants similarly. Other authors have reached similar conclusions in their studies where no statistically significant differences were found in the values obtained in both digital and physical models [20, 26–32].

Moreover, measurements in each arch were also analyzed, where software measurements were more reliable in the upper arch with no significant differences in the comparison of mean differences at T1 and T2 ($p=0.065$). In the lower arch, it seems that the measurements were reliable with all three methods.

Regarding the upper arch, significant differences were found in the measurements of the left and right sectors ($p=0.004$ and $p=0.001$ respectively) in the manual method, obtaining greater validity in the measurements by software and mobile app. And regarding the lower arch, significant differences were found in all sectors in the manual method (AAL: $p=0.022$, IAL: $p=0.010$, DAL: $p=0.013$), with no significant differences found in any of the sectors by the software method.

Therefore, it could be deduced that the software has been the method with the greatest reproducibility and temporal reliability and could be considered as an acceptable alternative in clinical practice. This result is consistent with various studies where the digital workflow using software demonstrated greater reproducibility and reliability, also showing that measurements were significantly faster to obtain compared to those made by the traditional manual method in plaster models [7, 13, 22]. According to different authors consulted, both measurements made by Orthocad [28, 33, 34], Emodels [7, 28], SureSmile [7, 34], AnatoModels [7], Mirror [7], Bibliocast Cécile 3 [33], and Maestro 3D [35] seem to be a good alternative in terms of accuracy and reliability compared to direct measurements on plaster models. This type of software can also be useful for performing other measurements in our field, such as analyzing and quantifying dental movement produced during treatment [36].

Furthermore, Pearson correlations were calculated between the measurements of each model (total and quadrant scores) at each time point. Although the correlations were significant, the measurements obtained from the mobile app "Ruler"; presented lower values than the other two models (0.414 for the quadrant correlation and 0.747 for the total correlation). According to the results obtained, the method with the least reliability and reproducibility would be the measurement by the mobile app "Ruler". This result is consistent with the study by RS Alwakeel et al. [5], where although the precision in the measurements of conventional models compared to those made using iPhone or Android software was similar, the measurements of available and required space using digitized photographs showed significant differences compared to the manual method. Other authors obtained different results. In the study by Prabhu K et al. [35] with the Scann 3D software for smartphones, although they observed, in the statistical analysis, that the standard deviation was greater in the premolar region, they were able to conclude that this mobile app was adequate and reliable compared to the traditional method; findings similar to the study by Morris RS et al. [37] where digital models created by the smartphone app DM were considered accurate enough for clinical use.

However, the results conducted so far on their reliability show contradictory results regarding the results obtained [5, 37]. Therefore, the use of mobile applications, in spatial analysis, still requires further scrutiny to demonstrate that they are useful and reliable for use as scientific and medical techniques.

One of the largest sources of random error according to different authors is the difficulty in identifying reference points [7, 19, 28, 30, 35]. In the present study, common definitions of anatomical points are adopted, considering that there will always be some imprecision in the identification of points, inherent to the analysis of models itself, regardless of the system used, i.e., a contact point could be a contact area, which would cause certain inaccuracies [19]. This task may be more difficult in digital measurements because examiners rely solely on subjective visual interpretations for the identification of reference points (since a three-dimensional structure appears as a two-dimensional image). Additionally, despite the high resolution of the software, the additional tactile sensation provided by physical models is absent with this new method. It should be noted that, in digital measurements, a specific limitation known as the "shape assumption" phenomenon occurs. The software inherently fills interproximal areas, when these are not entirely clear, are absent, or not captured, using patented mathematical algorithms, which is a potential source of inaccuracies. It may also be important to note that the operator's experience in making measurements may alter the results [15, 29].

These inaccuracies in measurements could also be due to another factor that alters the reliability or reproducibility of the values obtained, such as the training and experience of the examiner with the systems used in digital measurements [5, 19, 28, 30]. Reuschl RP et al. [28]. also suggest that there is a learning curve, more specifically in digital measurements, which will lead to smaller deviations as the observer becomes familiar with the technique. Therefore, errors are inevitable, as the measurements were carried out by human operators and these can improve with experience [30].

Tomasseti et al., analyzed the accuracy and efficiency of the total analysis using manual techniques and computerized methods and although they considered that all techniques are similar in terms of accuracy, the use of computerized methods could be more efficient in terms of the time used for measurements [14]. Other authors reached the same results [7, 28]. The comparative results improved in terms of reliability, validity, and time in the models analyzed digitally and obtained directly in the mouth by scanning instead of on the conventional plaster model [29, 38]. In the study carried out by Okamoto et al. [39], it was observed that the measurements made digitally by different operators were more consistent and reproducible than those made on a manual model.

No studies with similar objectives using artificial intelligence have been found in the literature. We consider that this aspect may be interesting for future studies.

One of the limitations when conducting our study was the difficulty in obtaining a larger homogeneous sample size in the study sample. Although the size used is similar to that of other studies analyzed [4, 7, 17, 28], we consider that an increase in the number of participants could help us improve our results. In the literature, we find that different authors analyze in their studies only one or two types of software in comparison with manual measurements, so we consider that, apart from increasing the sample size, adding more types of measurement software to the analysis would give us a more global view on the subject. Likewise, most of these studies, like ours, analyze permanent teeth and patients without growth, not finding any study that analyzes whether these results are comparable to those we would obtain if we studied deciduous dentition. We believe this could be a good future line of research.

## Conclusions

1. After the analysis carried out, it can be stated that the three methods used (manual method, Nemocast software and "Ruler" mobile application) obtained similar results with respect to the measurements

made. However, only the Nemocast software presented consistency in the measurements in both. temporal moments analyzed, being able to conclude that this measurement method would be the one with the greatest reproducibility and temporal reliability.

2. Although no significant differences were found both at the dental level and by quadrants, the digital method, specifically the Nemocast software, obtained greater validity and precision in the measurements of both arches with respect to the conventional method.
3. The use of new technologies for measuring models, especially the Nemocast 3D software, can facilitate our diagnostic work as well as its reproducibility and communication between professionals.

**Acknowledgements**
We would like to acknowledge the participation of all the patients.

**Author contributions**
A.A-L., L.A-Z., A.C. and D.G. conceived and designed the study; Data collection was conducted by A.A-L. L.A-Z. and A.C.; Analyses were performed by L.A.-Z. and A.A-L.; D.G., L.C-P., L.A-Z. and A.A-L. helped with data interpretation. A.A-L. and L.A-Z. led the writing. All authors have read and agreed to the published version of the manuscript.

**Funding**
This research received no external funding.

**Data availability**
Data is provided within the manuscript.

## Declarations

**Ethics approval and consent to participate**
All participants were informed about protocols, and they gave their written informed consent to participate in the study. The study followed the established guidelines by the Declaration of Helsinki for medical research involving human subjects and was approved by the bioethics committee of the University of Salamanca (USAL_23/1074).

**Consent for publication**
not applicable.

**Competing interests**
The authors declare no competing interests.

Received: 15 July 2024 / Accepted: 26 December 2024
Published online: 30 December 2024

**References**

1. Park S-H, Byun S-H, Oh S-H, Lee H-L, Kim J-W, Yang B-E, et al. Evaluation of the reliability, reproducibility, and validity of digital orthodontic measurements based on various digital models among young patients. J Clin Med. 2020;9(9):2728.
2. Cunha T de MA da, da S Barbosa I, Palma KK. Orthodontic digital workflow: devices and clinical applications. Dent Press J Orthod. 2021;26(6):e21spe6.
3. Pachêco-Pereira C, De Luca Canto G, Major PW, Flores-Mir C. Variation of orthodontic treatment decision-making based on dental model type: a systematic review. Angle Orthod. 2015;85(3):501–9.
4. Regragui S, Dakkaa A, Regraguib F, Bouklouzc A, Zaouia F. Which method to measure dentomaxillary discrepancy? Int Orthod. 2010;8:64–77.
5. Sami R, Awawdeh M, Barakah DM, Alwakeel RS. Measurement accuracy of space analysis by smart phone applications for orthodontic purposes: a comparison study with conventional plaster dental models. Int J Appl Dent Sci. 2018;4(3):314–22.
6. Ferreira JB, Christovam IO, Alencar DS, da Motta AFJ, Mattos CT, Cury-Saramago A. Accuracy and reproducibility of dental measurements on tomographic digital models: a systematic review and meta-analysis. Dentomaxillofac Radiol. 2017;46(7):20160455.
7. Grünheid T, Patel N, De Felippe NL, Wey A, Gaillard PR, Larson BE. Accuracy, reproducibility, and time efficiency of dental measurements using different technologies. Am J Orthod Dentofac Orthop. 2014;145(2):157–64.
8. Kumar AA, Phillip A, Kumar S, Rawat A, Priya S, Kumaran V. Digital model as an alternative to plaster model in assessment of space analysis. J Pharm Bioallied Sci. 2015;7(Suppl 2):S465–9.
9. Hayashi K, Chung O, Park S, Lee S-P, Sachdeva RCL, Mizoguchi I. Influence of standardization on the precision (reproducibility) of dental cast analysis with virtual 3-dimensional models. Am J Orthod Dentofac Orthop. 2015;147(3):373–80.
10. 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.
11. 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 Dentofac Orthop. 2016;149(2):161–70.
12. Leung C-YV, Yang Y, Liao C, Hägg U, Wong RWK, McGrath C, et al. Digital models as an alternative to plaster casts in assessment of orthodontic treatment outcomes. ScientificWorldJournal. 2018;2018:1–5.
13. Petrović V, Šlaj M, Buljan M, Čivljak T, Zulijani A, Perić B. Comparison of tooth size measurements in Orthodontics using conventional and 3D Digital Study models. J Clin Med. 2024;13(3):730.
14. Tomassetti JJ, Taloumis LJ, Denny JM, Fischer JR. A comparison of 3 computerized Bolton tooth-size analyses with a commonly used Method. Angle Orthod. 2001;71(5):351–7.
15. Garino F, Garino B. Comparison of Dental Arch measurements between Stone and Digital casts. W J O. 2002;3(N– 3):250–4.
16. Quimby ML, Vig KWL, Rashid RG, Firestone AR. The accuracy and reliability of measurements made on computer-based Digital models. Angle Orthod. 2004;74(3):298–303.
17. Nalcaci R, Topcuoglu T, Ozturk F. Comparison of Bolton analysis and tooth size measurements obtained using conventional and three-dimensional orthodontic models. Eur J Dent. 2013;7(1):S66–70.
18. Serrano-Velasco D, Martín-Vacas A, Giovannini G, Paz-Cortés MM, Aragoneses JM. Accuracy Analysis of Digital Models from Intraoral Scanners and 3D-Printed casts in Children and teenagers. Child (Basel). 2024;11(9):1082.
19. Koretsi V, Tingelhoff L, Proff P, Kirschneck C. Intra-observer reliability and agreement of manual and digital orthodontic model analysis. Eur J Orthod. 2018;40(1):52–7.
20. Gül Amuk N, Karsli E, Kurt G. Comparison of dental measurements between conventional plaster models, digital models obtained by impression scanning and plaster model scanning. Int Orthod. 2019;17(1):151–8.
21. Siddiqui NR, Hodges SJ, Sharif MO. Orthodontic apps: an assessment of quality (using the Mobile App Rating Scale (MARS)) and behaviour change techniques (BCTs). Prog Orthod. 2021;22(1):25.
22. Nandra S, Crawford N, Burford D, Pandis N, Cobourne MT, Seehra J. An investigation into the reliability of a mobile app designed to assess orthodontic treatment need and severity. Br Dent J. 2022;232(10):721–6.
23. Pojda D, Tomaka AA, Luchowski L, Tarnawski M. Integration and application of multimodal measurement techniques: relevance of photogrammetry to orthodontics. Sens (Basel). 2021;21(23):8026.
24. Moylan HB, Carrico CK, Lindauer SJ, Tüfekçi E. Accuracy of a smartphone-based orthodontic treatment-monitoring application: a pilot study. Angle Orthod. 2019;89(5):727–33.
25. Al-Mashraqi AA, Alhammadi MS, Gadi AA, Altharawi RA, Zamim KAH, Halboub E. Accuracy and reproducibility of permanent dentitions and dental arch measurements: comparing three different digital models with a plaster study cast. Int J Comput Dent. 2021;24(4):353–62.
26. Alrasheed WA, Owayda AM, Hajeer MY, Khattab TZ, Almahdi WH. Validity and reliability of intraoral and plaster models' photographs in the assessment of Little's Irregularity Index, tooth size-arch length discrepancy, and Bolton's analysis. Cureus. 2022;14(3):e23067.

27. Gupta G, Vaid NR. The world of orthodontic apps. APOS Trends Orthod. 2017;7(73):73–9.
28. Reuschl RP, Heuer W, Stiesch M, Wenzel D, Dittmer MP. Reliability and validity of measurements on digital study models and plaster models. Eur J Orthod. 2016;38(1):22–6.
29. Shailendran A, Weir T, Freer E, Kerr B. Accuracy and reliability of tooth widths and Bolton ratios measured by ClinCheck Pro. Am J Orthod Dentofac Orthop. 2022;161(1):65–73.
30. Wan Hassan WN, Othman SA, Chan CS, Ahmad R, Ali SN, Abd Rohim A. Assessing agreement in measurements of orthodontic study models: Digital caliper on plaster models vs 3-dimensional software on models scanned by structured-light scanner. Am J Orthod Dentofac Orthop. 2016;150(5):886–95.
31. Kumar AA, Ananthakrishnan MG, Kumar S, Divakar G, Sekar S, Dharani S. Assessing the validity and reliability of tooth widths and Bolton ratios obtained from digital models and plaster models. J Pharm Bioallied Sci. 2022;14(Suppl 1):S148–51.
32. Fleming PS, Marinho V, Johal A. Orthodontic measurements on digital study models compared with plaster models: a systematic review: Digital models: a systematic review. Orthod Craniofac Res. 2011;14(1):1–16.
33. Watanabe-Kanno GA, Abrão J, Junior HM, Sánchez-Ayala A, Lagravère MO. Determination of tooth-size discrepancy and Bolton ratios using bibliocast Cécile3 digital models. Int Orthod. 2010;8(3):215–26.
34. Westerlund A, Tancredi W, Ransjö M, Bresin A, Psonis S, Torgersson O. Digital casts in orthodontics: a comparison of 4 software systems. Am J Orthod Dentofac Orthop. 2015;147(4):509–16.
35. Prabhu K, Venkatesan N, Kirubakaran A, Karthick VC, Imithiyas M, Karthick R. Comparison of reliability and validity of measurements on digital study models made with scann three-dimensional smartphone software and plaster models: in vitro study. J Interdiscip Dent. 2021;11(1):11.
36. Nalcaci R, Kocoglu-Altan A, Bilcakci A, Oxturk F, Babacan H. A reliable method for evaluating upper molar distalization: superimposition of three-dimensional digital models. Korean J Orthod. 2015;45(2):82–8.
37. Morris RS, Hoye LN, Elnagar MH, Atsawasuwan P, Galang-Boquiren MT, Caplin J, et al. Accuracy of Dental Monitoring 3D digital dental models using photograph and video mode. Am J Orthod Dentofac Orthop. 2019;156(3):420–8.
38. Verma RK, Singh SP, Verma S, Kumar V, Bhupali NR, Arora S. Comparison of reliability, validity, and accuracy of linear measurements made on pre- and posttreatment digital study models with conventional plaster study models. J Orthodont Sci. 2019;8:18.
39. Okamoto A, Karibe H, Tanaka S, Kawakami T, Shinya A. Reliability of mixed dentition space analysis using a digital model obtained from an optical impression: a preliminary study. BMC Res Notes. 2024;17(1):12.

## Publisher's note