Reproducibility and Validity of Tooth Measurements Made from a 3D Digital Model Scanner Compared to Plaster Models

Scott Cardall, D.M.D.

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Orthodontics

Oregon Health & Science University
Portland, Oregon

December, 2012

Reproducibility and Validity of Tooth Measurements Made Using a 3D Digital Model Scanner Compared to Plaster Models

Scott Cardall, D.M.D.

Master of Science in Orthodontics Research Advisory Committee:

Signature: [redacted] Date: 12-13-12

Jennifer Crowe, D.D.S., M.S.
Assistant Professor
Department of Orthodontics
Oregon Health and Science University

Signature: [redacted] Date: 12/13/12

David A. Covell, Jr., Ph.D., D.D.S.
Associate Professor, Chair
Department of Orthodontics
Oregon Health and Science University

Signature: [redacted] Date: 12/13/2012

Larry Doyle, D.D.S.
Assistant Professor, Program Director
Department of Orthodontics
Oregon Health and Science University

Signature: [redacted] Date: 13 Dec 2012

Mark E. Engelstad, D.D.S., M.D., M.H.I.
Associate Professor of OMFS, Residency Program Director
Department of Oral & Maxillofacial Surgery
Oregon Health and Science University

Acknowledgements

Mansen Wang, Ph.D., M.S.
Ross Kaplan, B.D.S., M.S.D.
David May, D.M.D., Ph.D.
Daniel Egbert
Cathrine Martell
Shannon Woods
Brent Erickson
Kevin Andrews
Adam Ward

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Reproducibility and Validity of Tooth Measurements Made from a 3D Digital Model Scanner Compared to Plaster Models

Scott R. Cardall[a], Jennifer J. Crowe[a], David A. Covell, Jr.[a], Larry M. Doyle[a], Mark E. Engelstad[b]

[a]Department of Orthodontics, School of Dentistry, Oregon Health & Science University.

[b]Department of Oral & Maxillofacial Surgery, School of Dentistry, Oregon Health & Science University.

Key Words: lasers; imaging, three-dimensional; dental models; reproducibility of results; software validation; orthodontics

Corresponding Author:

Jennifer Crowe
Assistant Professor
Department of Orthodontics
Oregon Health & Science University
School of Dentistry SD24
611 SW Campus Drive
Portland, OR 97239
Telephone: 503-494-8921
Fax: 503-494-5777
crowej@ohsu.edu

Declaration of Interests

All authors express no conflict of interest in the study presented, financial or otherwise.

ABSTRACT

Objective: To determine the reproducibility and validity of automatically-determined and user-determined virtual dental model tooth width measurements, and to compare virtual dental model measurements to direct caliper measurements on plaster models.

Materials and Methods: Tooth widths and Bolton discrepancies were collected from forty-eight plaster models (24 maxillary, 24 mandibular); each model was measured two or more times using each of the three methods. Crowding was assessed by comparing models with mild (0-4.5 mm), moderate (4.6-9 mm), and severe crowding (> 9 mm; n=16 of each). Descriptive statistics were calculated, and method error, intra-rater reliability, and inter-rater reliability were evaluated using Dahlberg's formula, paired t-tests, Pearson's Correlation Coefficient, and the Concordance Correlation coefficient. Analysis of Variance (ANOVA) and Bland-Altman plots assessed inter-method effects.

Results: Measurement standard deviations averaged 0.12 mm for automatic virtual models, 0.16 mm for user-defined virtual models and 0.20 mm for plaster models. Mean differences between methods ranged from 0.04 mm to 0.17 mm. By tooth type, molar measurements had the highest variance (0.20 mm - 0.33 mm) and Bolton 6-6 measurements had the highest overall variance (0.59 mm to 1.59 mm). Pearson correlation coefficients were ≥0.999 and concordance correlation coefficients were ≥0.998 for each method regardless of crowding. Repeated-measures ANOVA found no significant differences by measurement method, arch, or crowding group.

Conclusions: Automatic and user-defined tooth measurements from virtual models have similar clinical acceptability relative to each other and to plaster model measurements.

INTRODUCTION

Evolving capabilities of computers have led to many advances in digital technology in dentistry including the creation and manipulation of virtual dental models.[1] Virtual models offer many benefits including decreased need for laboratory work, reduced physical storage space needs, increased accessibility, and increased convenience in exploring diagnostic set-ups.[2,3] However, despite general acceptance of digital photos, radiographs, and cephalometric tracing programs, the substitution of virtual models for plaster casts has not been widespread. Plaster models remain in common use and are viewed as the traditional standard.[2,3,4,5,6,7]

Beyond costs associated with 3D scanner technology, slow acceptance of digital models can be attributed to varying reports of reliability and reproducibility among scanners. In 1999, OrthoCad (Cadent, Carlstadt, NJ, USA) began offering the first widely available means of obtaining virtual models at a centralized facility. The approach involves multiple laser scans of a plaster model cut into thin slices (a process called destructive scanning) and virtually reconstructing the model with proprietary software.[3] This method is reproducible with high interrater reliability of measurements made from the digital models.[8,9,10] However, caliper measurements of plaster models have shown higher regression coefficient values and smaller inter- and intraobserver errors, suggesting better accuracy and reproducibility of plaster models.[9] Statistically significant discrepancies of 0.15-2.88 mm reported in OrthoCad virtual models,[11] and with variation between maxillary and mandibular arches.[8] Another study showed virtual model measurements were uniformly smaller than plaster measurements.[10] A second early option for digital models came from emodel (GeoDigm, Chanhassen, MN), which uses a nondestructive surface laser scanner at a centralized facility to virtually slice a plaster model and then reconstruct a virtual model using software.[3] Significant differences have been found between emodel virtual models and plaster models, including a study which found emodel virtual model measurements were slightly larger than plaster model measurements, although none of the differences were deemed clinically significant.[12,13] Scanners not specifically adapted for dentistry have also been evaluated. Photostereometric 3D images and dental casts have shown differences within the range of operator error at 0.27 mm.[14] Similarly, scanners using holographic sensors found a maximum mean difference of 0.28 mm,[15] while scanners using other line lasers have shown differences of 0.30 mm and 0.39 mm.[16,17] All of these values were considered clinically insignificant.

More recently, in-office surface laser scanners have become available. Investigation of the 3Shape D-250 scanner (Copenhagen, Denmark) found the reproducibility and accuracy of linear measurements made on virtual models were comparable to plaster cast measurements.[18] However, measuring requires a process called segmenting, which requires the operator to select various points for each tooth for the software to construct an estimate of the long axis of the tooth and tooth margin. The Ortho Insight 3D scanner (Motion View Software, Hixson, Tenn) uses a simplified measuring process and was built specifically for orthodontic use. Among other features, the software can measure tooth widths using either automatically-determined or user-determined mesiodistal contact points. Specialty-specific software and the simplicity of tooth measurement prompted our study of this in-office scanner.

The purpose of this study was to assess the reproducibility and validity of measurements made using Ortho Insight 3D virtual models relative to plaster models.

Because the Ortho Insight 3D software allows for either automatic detection or user selection of mesiodistal tooth dimensions, we compared both of these measurement methods to each other as well as to direct caliper measurements. Because past studies have noted concerns about the reliability of severely crowded model measurements,[9,19,20] mild, moderate, and severely crowded models were tested for reliability and reproducibility. We also compared maxillary and mandibular measurements because previous studies have found significant interarch differences.[8]

MATERIALS AND METHODS

Sample Selection

Approval for this study was obtained from the Oregon Health & Science University Institutional Review Board. To determine the sample size, a pilot study was performed using 18 plaster dental arch models. Data indicated that for a power ($1-\beta$) of 0.8, 38 models would be required to detect a measurement difference of <0.2 mm, similar to reliability thresholds set in similar studies,[7] and an amount considered the smallest tooth-size measurement that could affect treatment decisions. To maintain an adequate sample of each crowding category, the final sample was expanded to a total of 48 models (24 maxillary and 24 mandibular).

Pretreatment orthodontic plaster models were randomly selected from the Oregon Health & Science University orthodontic clinic's archives. Inclusion criteria required each model have: (1) full eruption of permanent dentition mesial to the second molar in each arch, (2) intact restorations (if present), and (3) no fractured, missing, or abnormally-shaped teeth, or voids in teeth.

Crowding Assessment

To assess for the effect of crowding on measurement reliability and reproducibility, 16 models each (eight maxillary and eight mandibular) were selected that contained mild (≤4.5 mm), moderate (4.6-9.0 mm), or severe (≥9.1 mm) crowding. Selection of the three groups was made by three orthodontists, each with at least 35 years of practice experience in arch length analysis, which involved adding first molar to first molar mesiodistal tooth widths and comparing this value to needed arch perimeter. The crowding assessments were then averaged to determine crowding group assignment.

Plaster Model Measurement

The examiners independently measured the teeth of the 48 models using digital calipers to the nearest .01-mm value (Pittsburgh 4" Electronic Digital Pointed-Jaw Caliper, Masel, Carlsbad, Calif). All examiners were trained for standardization. Multiple examiners started the measurement process, but because of time commitment required, only seven finished all measurements and were included in the study. The mesiodistal width of each tooth was defined and measured according to established methods as the greatest distance between ideal contact points. [2,11,12,21,22,23,24] At least 30 days after measuring the last model, each examiner remeasured six models—three maxillary and mandibular arches, one model for each crowding category. Full arch and anterior Bolton discrepancies were calculated from individual tooth size measurements.[22]

Automatic Virtual Model Measurement

One examiner (S.C.) scanned all models five times with the Ortho Insight 3D scanner using the default white plaster model scan exposure setting of 3.00 milliseconds. To determine proper scan resolution, a pilot study comparing high and low resolution settings revealed a relatively large measurement discrepancy of >0.5 mm on five of 28 teeth. In addition, visual ability to differentiate teeth was noticeably easier using the high resolution setting, so high resolution scans were used for the study.

The scanner software requires each tooth to be identified prior to measuring. For most teeth, the examiner placed a single point on the center of the clinical crown.

Occasionally, one or two other points around the margin of the clinical crown were needed to help the software identify the entire crown. The software then estimated the mesial and distal contact points of each tooth and used the points to determine tooth widths.

User-defined Virtual Model Measurement

For user-defined virtual model measurements, the same examiner scanned all models once more. Two examiners completed all user-defined tooth measurements. They identified each tooth as had been done for automatic virtual model measurements, then made user-defined mesial and distal contact points using the software features for visualization. These features included the ability to toggle between single tooth and full arch views, rotate teeth in three dimensions, zoom in and out, manipulate the vertical axis of the tooth, and use parallel vertical planes to discern the contact points (Figure 1). The software then measured tooth widths based on the points set by the examiner.

Calibration

To correlate scanner and caliper measurements, a uniform calibration block of 14.5 mm, a size near the maximum molar width, was fabricated using dental plaster. After the block was scanned and measured with calipers, two layers of dye stone hardener (Tanaka Stone Surface Sealer for Plaster, Skokie, Ill) were applied to resist wear during the measuring process with calipers. The block was then rescanned and remeasured with calipers to ensure the dye stone hardener did not change the dimension of the calibration block. To ensure caliper consistency, all seven examiners were instructed to zero the caliper with the jaw blades together and remeasure the calibration stone before measuring each model. At the end of the study, the calibration block was rescanned and remeasured with calipers to assess for wear.

Statistical Analysis

All statistical analyses were run using SAS version 9.3 (SAS Institute Inc, Cary, NC). Variables were assessed for normal distribution using the Kolmogorov-Smirnov test. Means and standard deviations for each tooth and method were calculated, as well as mean differences and standard deviation differences between each tooth and method. The means were then compared using paired t-tests, and a Tukey adjustment was performed to limit false positives resulting from multiple comparisons. For each examiner and each automatic virtual model scan, paired t-tests calculated intra-rater reliability and Dahlberg's Formula assessed method error. Pearson's correlation coefficient and concordance correlation coefficient (CCC) modified for multiple measurements was used to assess inter-rater reliability, while paired t-tests relative to mean values assessed significance. Repeated-measures analysis of variance (ANOVA) was used to detect inter-method significant differences between crowding groups (mild, moderate, and severe), arches (maxillary vs. mandibular), and measurement method (plaster model vs. automatic virtual model vs. user-defined virtual model). Bland-Altman plots with limits of agreement were used for visual assessment of measurement agreement.[25]

RESULTS

Direct Plaster Model Measurements

Inter-rater reliability as assessed with Pearson's Correlation Coefficient was high (>0.999) among the seven examiners. Reproducibility of plaster model measurements using the concordance correlation coefficient was 0.998±0.001 (95% CI) or higher for all 24 sets of models. Intra-rater paired t-tests and inter-rater paired t-tests relative to mean measurements found significant differences for several examiners (Table 4). Dahlberg values ranged from 0.260 to 0.796. Across all types of teeth, mesial-distal measurements had an average standard deviation of 0.21 mm. Among tooth type, incisor measurements had the lowest variability (standard deviation: 0.13 mm) and molar measurements had the highest (standard deviation: 0.33 mm; Figure 2). Anterior and posterior Bolton discrepancy calculations had standard deviations of 0.63 mm and 1.59 mm, respectively (Figure 2). Standard deviations for maxillary and mandibular teeth were not significantly different (Table 1).

Variance by crowding category revealed standard deviations of 0.19 mm for severe, 0.20 mm for moderate, and 0.23 mm for mild crowding, differences that were not significant (Table 2). No significant intra-arch or inter-arch differences were found when controlling for crowding (Table 2).

Automatic Virtual Model Measurement

Inter-rater reliability assessed with Pearson's Correlation Coefficient was 0.999 to 1.000 between the five automatically-measured virtual model scans. Reproducibility of automatically-measured virtual models using CCC was also 0.999-1.000 for all 24 sets of models. While intra-rater paired t-tests found no significant differences within scans, inter-rater paired t-tests relative to mean measurements found significant differences for four of five scans (Table 4). Dahlberg values ranged from 0.260 to 0.796. Across all types of teeth, mesial-distal measurements had an average standard deviation of 0.12 mm. Among tooth type, incisor measurements had the lowest variability (standard deviation: 0.09 mm) and molar measurements had the highest (standard deviation: 0.20 mm; Figure 2). Anterior and posterior Bolton discrepancy calculations had standard deviations of 0.35 mm and 0.59 mm, respectively (Figure 2). Standard deviations for maxillary and mandibular teeth were not significantly different (Table 1).

Variance by crowding category revealed standard deviations of 0.13 mm for severe, 0.12 mm for moderate, and 0.10 mm for mild crowding, differences that were not significant (Table 2). No significant intra-arch or inter-arch differences were found when controlling for crowding (Table 2).

User-defined Virtual Model Measurement

Inter-rater reliability with Pearson's Correlation Coefficient was 0.999 between the two examiners. Reproducibility of user-defined virtual model measurements using CCC was 0.999±0.001 (95% CI) or higher for all 24 sets of models. While intra-rater paired t-tests found no significant differences within examiner measurements, inter-rater paired t-tests found differences between examiners (Table 4). Dahlberg values were 0.254 and 0.387. Across all types of teeth, mesial-distal measurements had an average standard deviation of 0.19 mm. Among tooth type, canines had the lowest average measurement standard deviation (0.12 mm), with the greatest variability in molar measurements (0.24 mm; Figure 2). Anterior and posterior Bolton discrepancy calculations had standard deviations of 0.54 mm and 1.38 mm, respectively (Figure 2). Standard deviations for maxillary and mandibular teeth were not significantly different (Table 1).

Variance in measurements by amount of crowding present revealed standard deviations of 0.13 mm for high crowding, 0.12 mm for moderate crowding, and 0.10 mm for mild crowding, differences which were not significant (Table 2). No significant intra-arch or inter-arch differences were found when controlling for crowding (Table 2).

Plaster vs Automatic Virtual vs User-defined Virtual Model Measurements

Kolmogorov-Smirnov tests revealed normal distributions for plaster model, automatic virtual model, and user-defined virtual model measurements. Repeated-measures ANOVA revealed the only significant difference was that mandibular teeth had lower measurement values than maxillary teeth ($p<.0001$; Table 3), thus there were no significant differences found between measurement methods (Table 5), or when controlling for crowding (Table 2).

Comparison of automatic virtual model measurements to plaster model measurements found automatic measurements were on average 0.13 mm larger per tooth, with all mean differences <0.40 mm (Table 5). Comparison of user-defined virtual model measurements to plaster model measurements found user-defined measurements were on average 0.17 mm larger per tooth, with all mean differences <0.24mm. Comparison of automatic virtual model measurements to user-defined virtual model measurements found user-defined measurements were on average 0.04 mm larger per tooth, with all mean differences ≤0.30 mm. Bland-Altman plots with limits of agreement visually depicted this information by tooth size (Figure 3a-3c).

Mean differences in Bolton discrepancies showed the greatest variance (Table 5). Bolton 6-6 measurements showed a mean difference >2 mm in three combinations: between maxillary plaster and automatically-measured virtual models, maxillary plaster and user-defined virtual models, and mandibular plaster and user-defined virtual models. Other mean differences in Bolton discrepancies were <1.2 mm.

Calibration

Calibration stone measurements with each set of model measurements were within 0.05 mm for all examiners using direct measurements and had the same range using scanner measurements.

DISCUSSION

Multiple studies have compared plaster model measurements to virtual model measurements.[8,9,10,11,12,13,18] This was the first study to assess automated virtual model measurements relative to user-defined virtual and plaster model measurements.[9,20,12] The Ortho Insight 3D software has algorithms for either automated measurement of tooth widths or user-defined mesial-distal measurement of each tooth. We found both types of virtual model measurements were slightly larger than direct plaster model measurements (Table 5). These results are similar to those of some studies,[13] which state examiners have difficulty placing points without adjacent teeth and tend to overestimate distances using virtual models. Another explanation could be that the thickness of the caliper and interference from adjacent teeth or plaster could produce smaller plaster model measurements, whereas the algorithm used by the scanner software did not have this limitation. However, other studies have found virtual model measurements to be slightly smaller, a difference attributed to shrinkage of alginate during transport,[10] 2-D line measurement deformation[20,26] and the difficulty of choosing correct points.[20] Whether tooth sizes tend to be uniformly smaller or larger, it has been argued that a generalized and uniform difference does not affect the diagnostic utility of a ratio measurement such as the Bolton analysis.[10]

Our study assessed reproducibility rather than accuracy. This was necessary because plaster model measurement contains its own method error and therefore cannot be upheld as the known value. In our study, standard deviations of plaster models were actually higher than virtual models, further questioning the use of plaster models as an accuracy standard in virtual model studies. As a result, comparisons of reproducibility were made in our study. However, since plaster models are still the traditional method, we were assessed the validity of virtual model measurements relative to the traditional plaster model by comparing differences between the methods. Our results showed differences of 0.04 to 0.17 mm between the three methods similar to previous studies in magnitude,[10,12,13,20] but also in lack of significance.[9,14,18,20]

Concerns of measurement reproducibility in the presence of crowding are widespread,[8,9,12,17,18,19,20] with some attributing statistical discrepancies to crowding and questioning whether virtual models should be used for crowded dentitions.[20] Past studies have either avoided measuring crowded models,[27] measured only severely crowded models,[18] defined crowding as greater than three millimeters,[19] or combined large ranges of crowding such as moderate to severe in one group >4 mm.[20] Stevens et al measured three broad groups of crowding (0-4 mm, 4-8 mm, and >8 mm) and found no significant differences, but had a sample size of three to six models per group and did not control for the arch containing the crowding.[12] One of the problems of measuring crowding in a study assessing measurement methods is the method chosen to determine the amount of crowding cannot also be a measurement method. For this reason, we employed visual arch length crowding analysis as assessed by several experienced orthodontists and averaged the results. This method of averaging the results of three trials and using it as a standard was similar to the method employed by Tomassetti, who used the average of three sets of measurements as a standard for comparison to other measurement methods.[27] Our model sample of mild, moderate, and severe crowding had a sample size of 16 per group (8 maxillary, 8 mandibular) and similarly found no significant measurement reproducibility differences (Table 2). This study also made distinctions in reproducibility by tooth type, finding the highest variability in molar measurements, as has been previously reported,[9] indicating molars are difficult to consistently measure. This is especially true for maxillary first molars, which often have rhomboidal shapes making consistency more difficult.

In assessing Bolton Analysis measurements, we set clinical significance at >1.5 mm, similar to Tomassetti and Proffit.[27,28] We found significant Bolton 6-6 intra-method standard deviations differences up to 1.67 mm (Table 5). These values were lower than a study on plaster models using calipers which found differences of up to 3.7 mm.[19] Inter-method Bolton 6-6 standard deviations in our study were as high as 2.26 mm, similar to a value of 1.84 mm reported in Tomassetti's inter-method study of the Bolton Analysis (Table 5).[27] These higher variances for Bolton discrepancy measurements relative to tooth size measurements are due to cumulative variability of included tooth measurements. It is worth noting that a discrepancy of this size may be even more significant in a crowded borderline extraction case.[19,20]

Studies of reproducibility and validity are useful because they allow for the evaluation of a new method relative to the traditional method. Repeated-measures ANOVA found no measurement differences by crowding group or measurement method. The only significant finding was that cumulative tooth size was larger in the maxillary arch than mandibular arch—a well-known fact in dentistry which is the basis of the Bolton analysis.[22,29]

For all methods, intra-rater reliability (reproducibility) of measurements for all model pairs was high, similar to other virtual model studies.[9,11,18] We used Dahlberg's formula to determine method error. In our study, these values were usually slightly larger than 0.2 mm, the minimum measurement difference detectable according to our power analysis, and a small percentage relative to measurement values (Table 4), with the exception of one examiner. With measurements as small at 4.5 mm for lower central incisors, a 0.796 mm method error equates to nearly 18% error. That one examiner had a higher method error suggests that examiners tasked with measuring plaster models should be screened for reproducibility.

Maxillary and mandibular teeth were also equally reproducible (Table 1). Consistency of calibration stone measurements suggests consistent calibration of measurements throughout the study. Consistency could have also been aided by methodology. Unlike destructive model scanning techniques, this study was able to avoid impression-related discrepancies between virtual and plaster models by using the same original models,[8,9,10] removing this variable from expected error.

Previous investigations of the validity and reproducibility of virtual models relative to plaster models have concluded virtual models are inferior (yet clinical acceptable),[9] offer questionable reliability in crowded cases,[20] or should garner general acceptance.[12,18] The results of this study indicate that both types of virtual model measurements from the Ortho Insight 3D scanner are reproducible and valid compared to plaster model measurements.

While the calibration process provided increased confidence in our measuring consistency between methods and within methods, a weakness of this study is that accuracy could not be assessed as no measurement method could be relied on to produce the true measurement value. We also assessed only the plaster model scanning feature of the Ortho Insight 3D scanner. Future research into the Ortho Insight 3D scanner could assess the reproducibility and validity of tooth size and Bolton measurements made from alginate impression scans.

CONCLUSIONS

1) Caliper-measured plaster models, automatically-measured virtual models, and user-defined virtual model measurements all provide reproducible and reliable measurements, with the possible exception of Bolton 6-6 measurements.

2) Reproducibility and reliability of the three measurement methods are not affected by which arch is measured or by the presence of mild, moderate, or severe crowding.

3) Intra-rater and inter-rater reliability is high for all three methods.

REFERENCES

1. Paredes, V. , J. L. Gandia , and R. Cibrian . Digital diagnosis records in orthodontics. An overview. *Med Oral Patol Oral Cir Bucal.* 2006;11:E88-93.
2. Dalstra, M. and B. Melsen . From alginate impressions to digital virtual models: accuracy and reproducibility. *J Orthod.* 2009;36:36-41.
3. Fleming, P. S. , V. Marinho , and A. Johal . Orthodontic measurements on digital study models compared with plaster models: a systematic review. *Orthod Craniofac Res.* 2011;14:1-16.
4. Rheude, B. , P. L. Sadowsky , A. Ferriera , and A. Jacobson . An evaluation of the use of digital study models in orthodontic diagnosis and treatment planning. *Angle Orthod.* 2005;75:300-304.
5. White, A. J. , D. W. Fallis , and K. S. Vandewalle . Analysis of intra-arch and interarch measurements from digital models with 2 impression materials and a modeling process based on cone-beam computed tomography. *Am J Orthod Dentofacial Orthop.* 2010;137:456.e1-9; discussion 456-7.
6. Motohashi, N. and T. Kuroda . A 3D computer-aided design system applied to diagnosis and treatment planning in orthodontics and orthognathic surgery. *Eur J Orthod.* 1999;21:263-274.
7. Schirmer, U. R. and W. A. Wiltshire . Manual and computer-aided space analysis: a comparative study. *Am J Orthod Dentofacial Orthop.* 1997;112:676-680.
8. Leifert, M. F. , M. M. Leifert , S. S. Efstratiadis , and T. J. Cangialosi . Comparison of space analysis evaluations with digital models and plaster dental casts. *Am J Orthod Dentofacial Orthop.* 2009;136:16.e1-4; discussion 16.

9. Zilberman, O. , J. A. Huggare , and K. A. Parikakis . Evaluation of the validity of tooth size and arch width measurements using conventional and three-dimensional virtual orthodontic models. *Angle Orthod.* 2003;73:301-306.

10. Santoro, M. , S. Galkin , M. Teredesai , O. F. Nicolay , and T. J. Cangialosi . Comparison of measurements made on digital and plaster models. *Am J Orthod Dentofacial Orthop.* 2003;124:101-105.

11. Quimby, M. L. , K. W. Vig , R. G. Rashid , and A. R. Firestone . The accuracy and reliability of measurements made on computer-based digital models. *Angle Orthod.* 2004;74:298-303.

12. Stevens, D. R. , C. Flores-Mir , B. Nebbe , D. W. Raboud , G. Heo , and P. W. Major . Validity, reliability, and reproducibility of plaster vs digital study models: comparison of peer assessment rating and Bolton analysis and their constituent measurements. *Am J Orthod Dentofacial Orthop.* 2006;129:794-803.

13. Mullen, S. R. , C. A. Martin , P. Ngan , and M. Gladwin . Accuracy of space analysis with emodels and plaster models. *Am J Orthod Dentofacial Orthop.* 2007;132:346-352.

14. Bell, A. , A. F. Ayoub , and P. Siebert . Assessment of the accuracy of a three-dimensional imaging system for archiving dental study models. *J Orthod.* 2003;30:219-223.

15. Ayoub, A. F. , D. Wray , K. F. Moos , J. Jin , T. B. Niblett , C. Urquhart , P. Mowforth , and P. Siebert . A three-dimensional imaging system for archiving dental study casts: a preliminary report. *Int J Adult Orthodon Orthognath Surg.* 1997;12:79-84.

16. Hirogaki, Y. , T. Sohmura , H. Satoh , J. Takahashi , and K. Takada . Complete 3-D reconstruction of dental cast shape using perceptual grouping. *IEEE Trans Med Imaging.* 2001;20:1093-1101.

17. Asquith, J. , T. Gillgrass , and P. Mossey . Three-dimensional imaging of orthodontic models: a pilot study. *Eur J Orthod.* 2007;29:517-522.

18. Sousa, M. V. , E. C. Vasconcelos , G. Janson , D. Garib , and A. Pinzan . Accuracy and reproducibility of 3-dimensional digital model measurements. *Am J Orthod Dentofacial Orthop.* 2012;142:269-273.

19. Shellhart, W. C. , D. W. Lange , G. T. Kluemper , E. P. Hicks , and A. L. Kaplan . Reliability of the Bolton tooth-size analysis when applied to crowded dentitions. *Angle Orthod.* 1995;65:327-334.

20. Redlich, M. , T. Weinstock , Y. Abed , R. Schneor , Y. Holdstein , and A. Fischer . A new system for scanning, measuring and analyzing dental casts based on a 3D holographic sensor. *Orthod Craniofac Res.* 2008;11:90-95.

21. Jensen, E. , P. K. Yen , C. F. Moorrees , and S. O. Thomsen . Mesiodistal crown diameters of the deciduous and permanent teeth in individuals. *J Dent Res.* 1957;36:39-47.

22. Bolton, W. A. Disharmony in tooth size and its relation to the analysis and treatment of malocclusion. *Angle Orthod.* 1958;28:113-130.

23. Nance, H. N. The limitations of orthodontic treatment; mixed dentition diagnosis and treatment. *Am J Orthod.* 1947;33:177-223.

24. Neff, C. W. Tailored occlusion with the anterior coefficient. *Am J Orthod.* 1949;35:309-313.

25. Myles, P. S. and J. Cui . Using the Bland-Altman method to measure agreement with repeated measures. *Br J Anaesth.* 2007;99:309-311.

26. Yen, C. H. Computer-aided space analysis. *J Clin Orthod.* 1991;25:236-238.

27. Tomassetti, J. J. , L. J. Taloumis , J. M. Denny , and J. R. Fischer Jr . A comparison of 3 computerized Bolton tooth-size analyses with a commonly used method. *Angle Orthod.* 2001;71:351-357.

28. Proffit W. R. Contemporary Orthodontics. 5th ed. St. Louis, MO Elsevier Mosby. 2012.

29. Gilpatrick, W. H. Arch Predetermination: Is it Practical? *Journal of the American Dental Association*. 1923;7:533.

Figure Legends:

Figure 1: Placing Mesiodistal Contact Points for User-defined Virtual Model Measurements

Figure 2: Variation Among Measurement Method by Tooth Grouping

Figure 3a: Bland-Altman Plot #1 - Automatically-measured Virtual Models vs User-determined Virtual Model Measurements

Figure 3b: Bland-Altman Plot #2 - Plaster Model Measurements vs User-determined Virtual Model Measurements

Figure 3c: Bland-Altman Plot #3 - Plaster Model Measurements vs Automatically-measured Virtual Models

Figure 1

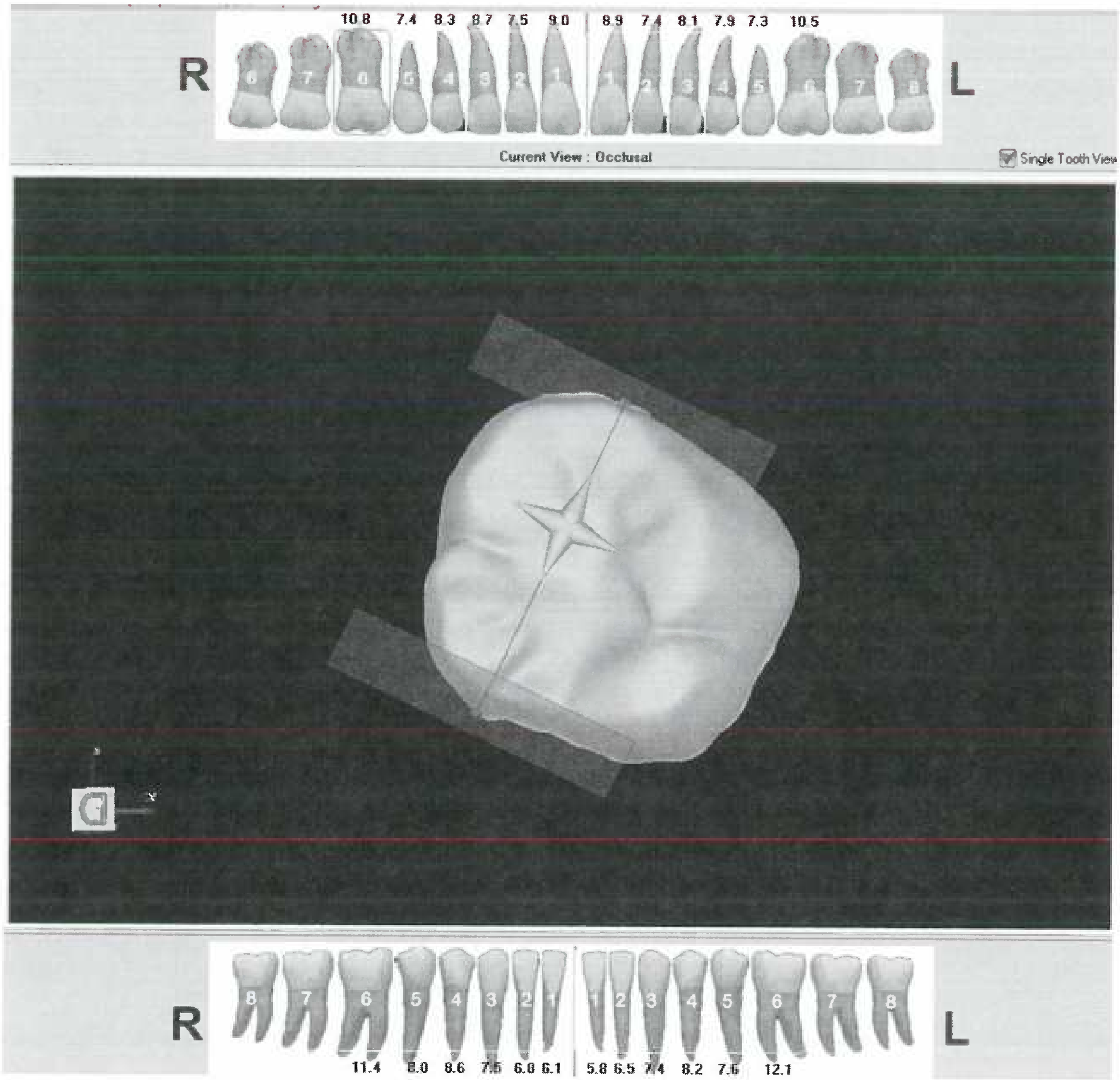

10.8 7.4 8.3 8.7 7.5 9.0 8.9 7.4 8.1 7.9 7.3 10.5
R
L
Current View : Occlusal
Single Tooth View
R
L
11.4 8.0 8.6 7.5 6.8 6.1 5.8 6.5 7.4 8.2 7.6 12.1

Figure 2

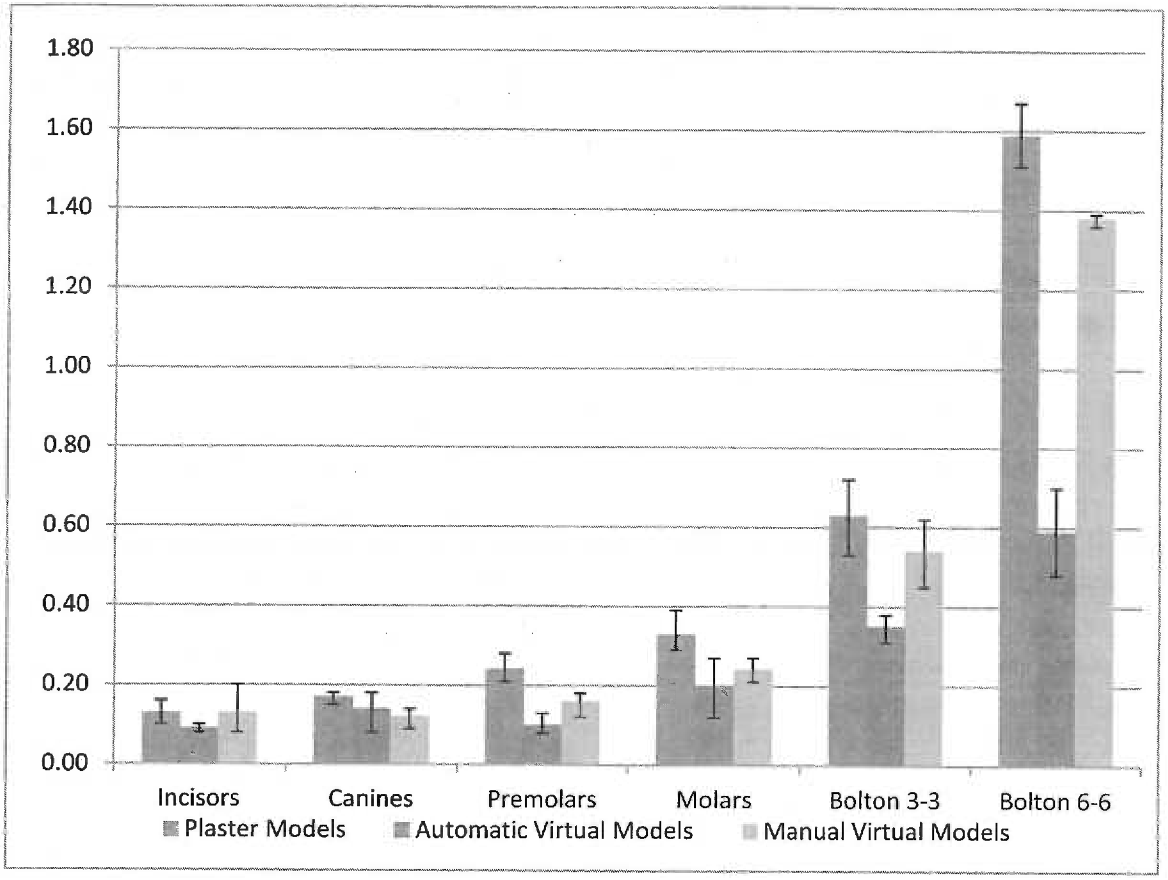

1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Incisors
Canines
Premolars
Molars
Bolton 3-3
Bolton 6-6
Plaster Models
Automatic Virtual Models
Manual Virtual Models

Figure 3a

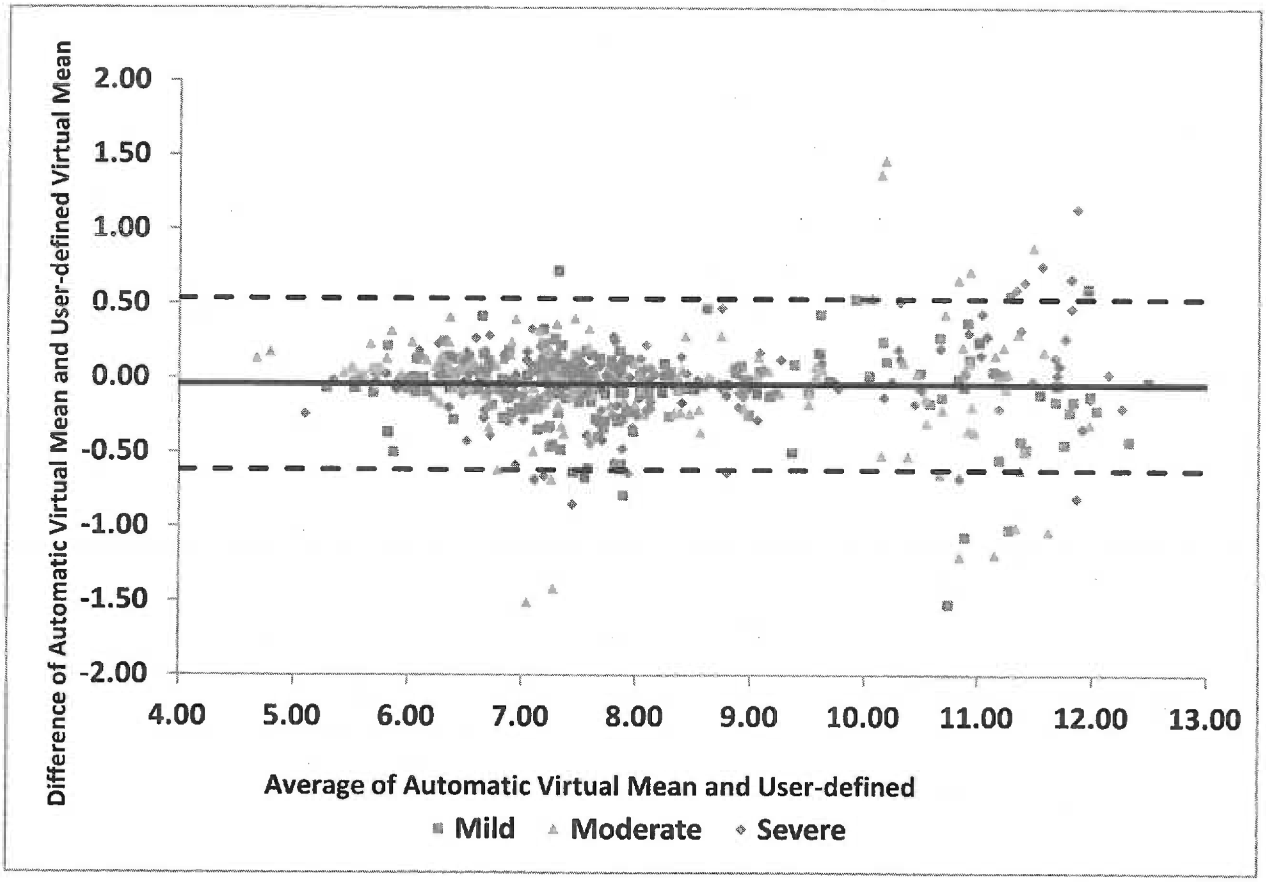
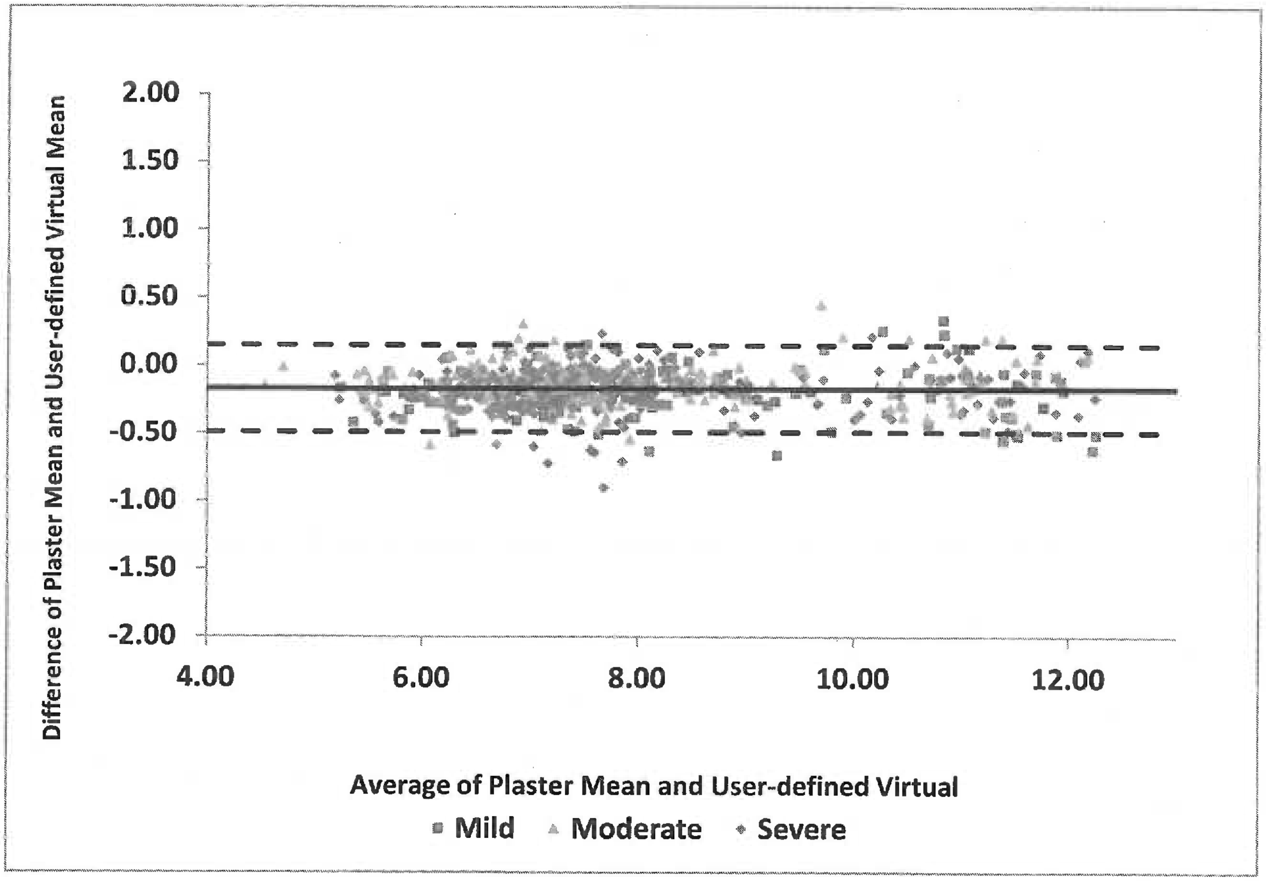

2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
11.00
12.00
13.00
Difference of Automatic Virtual Mean and User-defined Virtual Mean
Average of Automatic Virtual Mean and User-defined
Mild
Moderate
Severe

Figure 3b



2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
4.00
6.00
8.00
10.00
12.00
Difference of Plaster Mean and User-defined Virtual Mean
Average of Plaster Mean and User-defined Virtual
Mild
Moderate
Severe

Figure 3c

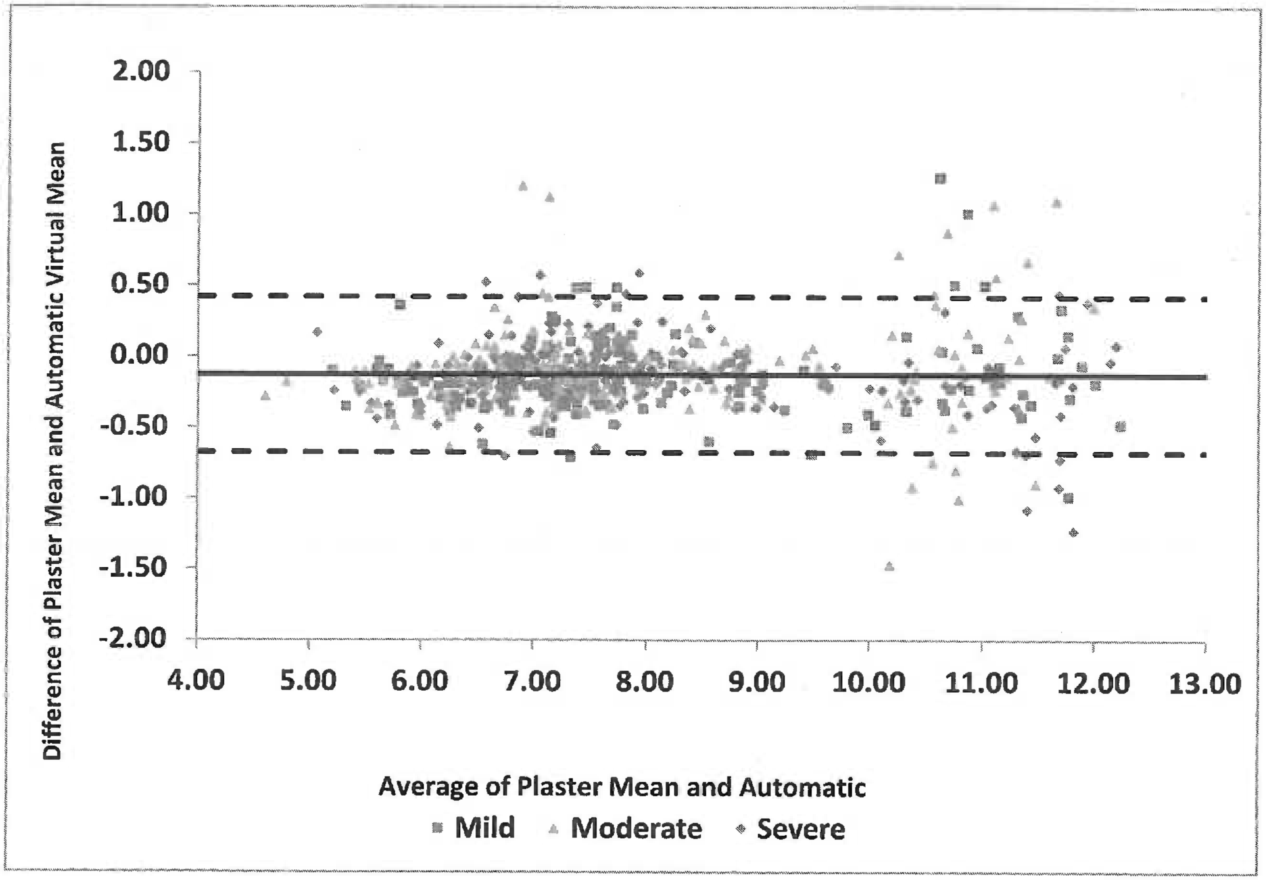

2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
Difference of Plaster Mean and Automatic Virtual Mean
4.00
5.00
6.00
7.00
8.00
9.00
10.00
11.00
12.00
13.00
Average of Plaster Mean and Automatic
Mild
Moderate
Severe

Table 1: Method Measurement Variation by Arch – Standard Deviations

	Plaster Models	Automatic Virtual Models	Manual Virtual Models
Maxillary Arch	0.21 mm	0.1 mm	0.16 mm
Mandibular Arch	0.1966 mm	0.14 mm	0.16 mm

Table 2: Method Measurement Variation by Amount of Crowding– Standard Deviations

	Plaster Models	Automatic Virtual Models	Manual Virtual Models	
Mild Crowding	0.23 mm	0.1 mm	0.13 mm	
Moderate Crowding	0.2 mm	0.12 mm	0.14 mm	
Severe Crowding	0.19 mm	0.13 mm	0.21 mm	

Table 3: Repeated-measures ANOVA – significance of measurements controlled by method, crowding, and arch

	P value
Measurement Method	0.3283
Crowding	0.8476
Measurement Method + Crowding	0.9643
Arch	<.0001*
Measurement Method + Arch	0.7776
Arch + Crowding	0.3919
Measurement Method + Crowding + Arch	0.9998

Table 4: Intra-rater reliability, Inter-rater reliability, and method error

Plaster Models				Automatic Virtual Models				Manual Virtual Models			
Examiner	Dahlberg	*Paired t-test (P-value)	**Paired t-test (P-value)	**Examiner (scan)**	Dahlberg	*Paired t-test (P-value)	**Paired t-test (P-value)	**Examiner**	Dahlberg	*Paired t-test (P-value)	***Paired t-test (P-value)
1	0.293	0.2157	0.0001	**1**	0.329	0.3827	0.0001	**1**	0.387	0.5476	0.0001
2	0.329	0.7768	0.0001	**2**	0.056	0.7846	0.0009	**2**	0.254	0.3979	0.0001
3	0.265	0.1977	0.7731	**3**	0.069	0.1158	0.0548				
4	0.476	0.0082	0.0001	**4**	0.341	0.1941	0.0001				
5	0.373	0.001	0.639	**5**	0.057	0.8204	0.0001				
6	0.260	0.0021	0.0003								
7	0.796	0.0014	0.0001								

Table 5a: Maxillary Mean Tooth Widths, Standard Deviations, and Differences between Measurement Method

Tooth / Bolton	Plaster Mean (mm)	Plaster SD (mm)	Automatic Mean (mm)	Automatic SD (mm)	Manual Mean (mm)	Manual SD (mm)	Difference Plaster vs Automatic Mean (mm)	Difference Plaster vs Automatic SD (mm)	Difference Plaster vs Manual Mean (mm)	Difference Plaster vs Manual SD (mm)	Difference Automatic vs Manual Mean (mm)	Difference Automatic vs Manual SD (mm)
UL1	9.042	0.122	9.179	0.090	9.220	0.110	-0.137	0.031	-0.178	0.012	-0.041	-0.020
UL2	7.194	0.159	7.324	0.089	7.410	0.160	-0.130	0.070	-0.216	-0.001	-0.086	-0.071
UL3	8.234	0.183	8.201	0.138	8.360	0.110	0.033	0.045	-0.126	0.073	-0.159	0.028
UL4	7.338	0.248	7.559	0.082	7.520	0.130	-0.220	0.166	-0.182	0.118	0.039	-0.048
UL5	6.960	0.269	7.135	0.085	7.190	0.120	-0.175	0.185	-0.230	0.149	-0.055	-0.035
UL6	10.665	0.346	11.050	0.163	10.820	0.230	-0.385	0.183	-0.155	0.116	0.230	-0.067
UR1	9.053	0.117	9.305	0.095	9.260	0.160	-0.253	0.023	-0.207	-0.043	0.045	-0.065
UR2	7.277	0.137	7.468	0.083	7.460	0.200	-0.191	0.053	-0.183	-0.063	0.008	-0.117
UR3	8.202	0.151	8.302	0.084	8.370	0.130	-0.099	0.067	-0.168	0.021	-0.068	-0.046
UR4	7.352	0.214	7.539	0.089	7.500	0.130	-0.187	0.125	-0.148	0.084	0.038	-0.041
UR5	7.035	0.234	7.161	0.083	7.180	0.180	-0.126	0.151	-0.145	0.054	-0.019	-0.097
UR6	10.759	0.388	11.153	0.122	10.850	0.210	-0.394	0.266	-0.091	0.178	0.303	-0.088

Table 5b: Mandibular Mean Tooth Widths, Standard Deviations, and Differences between Measurement Method

Tooth / Bolton	Plaster Mean (mm)	Plaster SD (mm)	Automatic Mean (mm)	Automatic SD (mm)	Manual Mean (mm)	Manual SD (mm)	Difference Plaster vs Automatic Mean (mm)	Difference Plaster vs Automatic SD (mm)	Difference Plaster vs Manual Mean (mm)	Difference Plaster vs Manual SD (mm)	Difference Automatic vs Manual Mean (mm)	Difference Automatic vs Manual SD (mm)
LL1	5.671	0.145	5.907	0.087	5.880	0.110	-0.236	0.057	-0.209	0.035	0.027	-0.023
LL2	6.318	0.102	6.556	0.087	6.500	0.120	-0.238	0.015	-0.182	-0.018	0.056	-0.033
LL3	7.051	0.174	7.034	0.183	7.260	0.140	0.017	-0.009	-0.209	0.034	-0.226	0.043
LL4	7.399	0.206	7.453	0.107	7.520	0.180	-0.055	0.099	-0.121	0.026	-0.067	-0.073
LL5	7.408	0.252	7.454	0.119	7.510	0.180	-0.046	0.133	-0.102	0.072	-0.056	-0.061
LL6	11.297	0.288	11.179	0.268	11.460	0.260	0.118	0.021	-0.163	0.028	-0.281	0.008
LR1	5.643	0.116	5.834	0.083	5.820	0.080	-0.190	0.033	-0.177	0.036	0.014	0.003
LR2	6.242	0.126	6.458	0.102	6.450	0.130	-0.217	0.024	-0.208	-0.004	0.008	-0.028
LR3	7.074	0.177	7.163	0.172	7.260	0.090	-0.089	0.005	-0.186	0.087	-0.097	0.082
LR4	7.371	0.209	7.426	0.093	7.520	0.160	-0.055	0.116	-0.149	0.049	-0.094	-0.067
LR5	7.387	0.279	7.429	0.128	7.610	0.230	-0.042	0.151	-0.223	0.049	-0.181	-0.102
LR6	11.286	0.285	11.171	0.243	11.460	0.270	0.114	0.042	-0.174	0.015	-0.289	-0.027

Table 5c: Mean Tooth Widths, Standard Deviations, Bolton Measurements, and Differences between Measurement Method

Tooth / Bolton	Plaster Mean (mm)	Plaster SD (mm)	Automatic Mean (mm)	Automatic SD (mm)	Manual Mean (mm)	Manual SD (mm)	Difference Plaster vs Automatic Mean (mm)	Difference Plaster vs Automatic SD (mm)	Difference Plaster vs Manual Mean (mm)	Difference Plaster vs Manual SD (mm)	Difference Automatic vs Manual Mean (mm)	Difference Automatic vs Manual SD (mm)
All teeth	7.886	0.205	8.018	0.120	8.058	0.159	-0.133	0.086	-0.172	0.046	-0.040	-0.039
MD3-3	37.994	0.723	38.953	0.378	39.170	0.450	-0.959	0.345	-1.176	0.273	-0.217	-0.072
MD6-6	90.148	1.669	91.067	0.699	92.250	1.390	-0.918	0.970	-2.102	0.279	-1.183	-0.691
MX3-3	48.999	0.532	49.758	0.314	50.070	0.620	-0.759	0.218	-1.071	-0.088	-0.312	-0.306
MX6-6	99.091	1.511	101.353	0.480	101.130	1.360	-2.261	1.031	-2.039	0.151	0.223	-0.880

[1] Paredes V, Gandia JL, Cibrian R. Digital diagnosis records in orthodontics: an overview. *Med Oral Patol Oral Cir Bucal* 2006; 11:E88-E93
[2] Dalstra M, Melsen B. From alginate impressions to digital virtual models: accuracy and reproducibility. *Journal of Orthodontics*, Vol 36, 2009, 36-41
[3] Fleming PS, Marinho V, Johal A. Orthodontic measurements on digital study models compared with plaster models: a systematic review. *Orthod Craniofac Res* 2011; 14:1-16
[4] Rheude B, Sadowsky PL, Ferriera A, Jacobson A. An evaluation of the use of digital study models in orthodontic diagnosis and treatment planning. Angle Orthod 2005; 75:300-4
[5] White, AJ, Fallis DW, Vandewalle KS. Analysis of intra-arch and interarch measurements from digital models with 2 impression materials and a modeling process based on cone-beam computed tomography. *Am J Orthod Dentofacial Orthop* 2010; 137:456.e1-456.e9
[6] Motohashi N, Kuroda T. A 3D computer-aided design system applied to diagnosis and treatment planning in orthodontics and orthognathic surgery. Eur J Orthod 1999;21:263–74.
[7] Schirmer UR, Wiltshire WA. Manual and computer aided space analysis: a comparative study. Am J Orthod Dentofacial Orthop 1997;112:676–80.
[8] Leifert MF, Leifert MM, Efstratiadis SS, Cangialosi TJ. Comparison of space analysis evaluations with digital models and plaster dental casts. Am J Orthod Dentofacial Orthop 2009;136:16.e1-4
[9] Zilberman O, Huggare JA, Parikakis KA. Evaluation of the validity of tooth size and arch width measurements using conventional and three-dimensional virtual orthodontic models. Angle Orthod 2003;73:301-6
[10] 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;124:101-5
[11] Quimby ML, Vig KWL, Rashid RG, Firestone AR. The accuracy and reliability of measurements made on computer-based digital models. *Angle Orthod* 2004; **74**: 298–303
[12] Stevens DR, Flores-Mir C, Nebbe B, Raboud DW, Heo G, Major PW. Validity, reliability, and reproducibility of plaster vs digital study models: comparison of peer assessment rating and Bolton analysis and their constituent measurements. *Am J Orthod Dentofacial Orthop* 2006; **129**: 794–803
[13] Mullen SR, Martin CA, Ngan P, Gladwin M. Accuracy of space analysis with emodels and plaster models. Am J Orthod Dentofacial Orthop 2007;132:346-52.
[14] Bell A, Ayoub AF, Siebert P. Assessment of the accuracy of a three-dimensional imaging systems for archiving dental study models. J Orthod 2003;30:219-23.
[15] Ayoub AF, Wray D, Moos KF, Jin J, Niblett TB, Urquhart C, *et al.* A three-dimensional imaging system for archiving dental study casts: a preliminary report. *Int J Adult Orthod Orthognath Surg* 1997; **12**: 79–84.
[16] Hirogaki Y, Sohmura T, Satoh H, Takahashi J, Takada K. Complete 3D reconstruction of dental cast shape using perceptual grouping. 1EEE Trans Med Imaging 2001; 20: 1093–1101
[17] Asquith J, Gillgrass T, Mossey P. Three-dimensional imaging of orthodontic models: a pilot study. Eur J Orthod 2007;29: 517-22.
[18] Sousa MV. Accuracy and reproducibility of 3-dimensional digital model measurements. Am J Orthod Dentofacial Orthop 142:269 (2012)
[19] Shellhart WC, Lange DW, Kluemper GT, Hicks EP, Kaplan AL. Reliability of the Bolton tooth-size analysis when applied to crowded dentitions. *Angle Orthod.* 1995;65:327–334.
[20] Redlich M, Weinstock T, Abed Y, Schneor R, Holdstein Y, Fischer A. A new system for scanning, measuring and analyzing dental casts based on a 3D holographic sensor. Orthod Craniofac Res 2008;11:90-5
[21] Jensen E Moorrees CFA. Thomsen S O, Yen P K 1957 Mesiodistal crown diameters of the deciduous and permanent teeth in individuals. *Journal of Dental Research* 36 : 39 – 47
[22] Bolton WA. Disharmony in tooth size and its relation to the analysis and treatment of malocclusion. *Angle Orthod.* 1958;28:113–130.
[23] Nance HN, The limitations of orthodontic treatment. I. Mixed Dentition Diagnosis and Treatment II. Diagnosis and Treatment in the Permanent Dentition. American Journal of Orthodontics and Oral Surgery 33:177-223, 253-301
[24] Neff CW. Tailored occlusion with the anterior coefficient. American Journal of Orthodontics 1949;35: 309–314
[25] P. S. Myles and J. Cui. Using the Bland–Altman method to measure agreement with repeated measures Br. J. Anaesth. (2007) 99(3): 309-311
[26] Chen-Hsing Y. Computer aided space analysis. J Clin Orthod. 1991;4:236–8.

[27] Tomassetti JJ, Taloumis LJ, Denny JM, Fischer JR, Jr. A comparison of 3 computerized bolton tooth-size analyses with a commonly used method. Angle Orthod. 2001;71:351–357.
[28] Proffit W. Contemporary orthodontics. St Louis: Mosby; 2012. p. 183-184.
[29] Gilpatrick WH. **Arch Predetermination. Is it Practical?.** *J. A. D. A.* 1923;7:533