Validity and Reliability of Peer Assessment Rating Index Measurement Derived from Digital and Plaster Models

Abstract

Aims: The aim of the study was to determine the validity and reliability of Peer Assessment Rating (PAR) index score derived from digital and plaster models of the same patient. **Subjects and Methods:** Thirty orthodontic plaster study models were digitalized using the 3Shape R700™ Orthodontic 3D scanner. PAR Index scoring was carried out on both the plaster and digital models by one independent examiner calibrated in the PAR Index. The measurements were repeated at a second sitting. Measurements were made on plaster models with the PAR Index ruler and on digital models with the 3Shape OrthoAnalyzer™ software. **Statistical Analysis Used:** Bland-Altman plots were used to test for validity and intraexaminer reliability. **Results:** For PAR Index score, overjet and overbite component scores, 28 out of 30 measurements were within 95% limits of agreement. Other components of the PAR Index score had all points within 95% limits of agreement. For intraexaminer reliability, digital models had 28 out of 30 measurements and plaster models had 27 out of 30 measurements that were within 95% limits of agreement. **Conclusions:** Digital models are a clinically acceptable alternative to plaster models in the measurement of the PAR Index. Improvement in software design is necessary to attain greater agreement in the measurement of the overjet and overbite components of the PAR index score between plaster and digital models.

Keywords: *Digital models, Peer Assessment Rating index, plaster models*

**Yew Jia Peh,
Ming Tak Chew,
Hung Chew Wong[1],
Mimi Yow**

*Department of Orthodontics,
National Dental Centre
Singapore, [1]Yong Loo Lin
School of Medicine, National
University of Singapore*

Introduction

Study models provide a three-dimensional record of the occlusion and are an essential component of orthodontic treatment records. Plaster models are accurate, dimensionally stable, easy to use, and cost-effective. However, their disadvantages include the need for storage space, are bulky, heavy, and fragile, and require manual retrieval and transportation from storage area to the clinical area.[1-4]

Replacement of plaster models with digital models can save space, storage costs, and eliminate transportation logistics. There is also ease in organizing, searching, and retrieval of the digitized records of each patient. As such, there is a growing trend of replacing plaster models with digital models.[5,6]

Peer Assessment Rating index

The Peer Assessment Rating (PAR) index, formulated in 1987, provides a single summary score for all the occlusal anomalies which may be found in a malocclusion. It is made up of a number of subcomponents, each measuring distinct occlusal traits. The upper and lower anterior segments and buccal occlusion are given a weightage of one. Overjet is given a weightage of six, overbite is given a weightage of two, while centre line is given a weightage of four.[7]

The PAR Index is used as a standardized quantitative objective method to assess malocclusion and evaluate treatment standards and outcome of treatment.[8-13] It has been shown to be valid and reliable using plaster study models.[7,8,14,15] It is currently a contractual requirement of orthodontic practice in the National Health Service in England and Wales and many other countries. In our center, the PAR Index is used for auditing and monitoring of orthodontic treatment standard.

With the replacement of plaster models with digital models, evaluation of orthodontic treatment success will be carried out using scanned digital models only. As such, PAR Index scoring needs to be validated on

How to cite this article: Peh YJ, Chew MT, Wong HC, Yow M. Validity and reliability of peer assessment rating index measurement derived from digital and plaster models. APOS Trends Orthod 2018;8:131-8.

Address for correspondence:
*Dr. Yew Jia Peh,
Blk 1 Toh Yi Drive #07-139,
Singapore 591501, Singapore.
E-mail: yewjia@yahoo.com*

the digital models. This study was carried out to assess the validity and reliability of PAR Index measurements derived from digital study models produced by scanning the plaster study models using 3Shape R700™ Orthodontic 3D scanner and analyzed with OrthoAnalyzer™ software. There are other studies assessing alternative systems available worldwide,[16,17] but as this system is the one we are using, it was necessary to perform the task of validating this system independently.

Aims and objectives

The specific aims were to compare
1. Validity of the total PAR Index score derived from digital models compared to plaster models
2. Component scores of the PAR Index generated from plaster and digital models
3. Intraexaminer reliability of total PAR Index score derived from plaster and digital models.

Subjects and Methods

The research protocol was submitted to the Institutional Review Board (IRB) for ethical approval. IRB approval was obtained (CIRB reference code: 2013/426/D).

Thirty sets of pretreatment models, available in both digital and physical forms, were randomly selected.

The study models selected had full complement of permanent teeth with no retained deciduous teeth. Teeth present were grossly intact, with no large cavitations or restorations affecting tooth size and morphology. The computerized articulation of the digital models corresponded with the occlusion and articulation of the plaster models.

Study models were excluded when they include primary dentition, edentulous ridges due to missing teeth, features that altered the natural mesiodistal or buccolingual crown diameter, such as restorations, caries, attrition, and fracture, as well as plaster defects. Those with errors in computerized articulation of the digital models were also excluded from the study.

Material used for the plaster models is the American Dental Association Type II dental stone (WhipMix Corp, Louisville, Kentucky, USA).

The plaster models were scanned using the 3Shape R700™ Orthodontic 3D scanner [Figures 1-3] into digital models.

The same plaster models were used for physical measurements and were scored with the PAR Index ruler (Ortho-Care UK Ltd., copyright of the University of Manchester, UK), as described by Richmond *et al.*[7] The digital models were scored using the 3Shape OrthoAnalyzer™ (3Shape A/S, Copenhagen, Denmark) software.

PAR Index scoring was carried out on both the plaster and digital models by one independent examiner calibrated

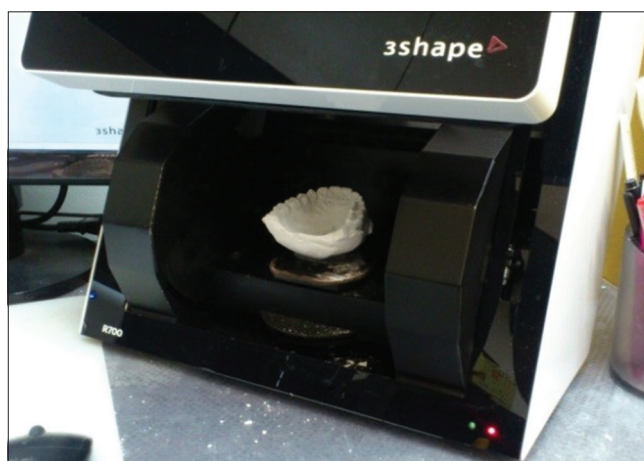**Figure 1: 3Shape R700™ orthodontic 3D scanner**

Figure 2: Scan of upper model

Figure 3: Scan of lower model

in the PAR Index. The examiner, who is the principal investigator, successfully completed a PAR Index calibration course held in Cardiff, UK, in February 2013. The examiner has also been trained in the usage of 3Shape

OrthoAnalyzer™ (3Shape A/S, Copenhagen, Denmark) software. PAR Index scoring for all 30 models was repeated at a second seating (T2) 4 weeks after the initial measurement.

Bland-Altman plots were used to test the validity and intraexaminer reliability.

Results

The PAR scores for the 30 sets of models are listed in Table 1.

Intraexaminer reliability

The Bland-Altman plots were used to check for intraexaminer reliability. Agreement is obtained if more than 95% of the points fall within the 95% limits of agreement.

The results for intraexaminer reliability [Table 3] showed that for plaster models, 27 out of 30 measurements were within the two standard deviations from the mean

difference. For digital models, 28 out of 30 measurements were within the two standard deviations from the mean difference. As <95% of the points were within the 95% limits of agreement, there is a lack of agreement in intraexaminer reliability for both plaster and digital models in the calculation of the total PAR Index score.

Physical measurements

From the Bland-Altman plot [Figure 4], out of 30 measurements, three observations were beyond the range of −4.07 and 4.40. There is a lack of agreement as <95% of the points lie within the 95% limits of agreement.

Digital measurements

From the Bland-Altman plot [Figure 5], out of 30 measurements, two observations were beyond the range of −2.45 and 2.58. There is a lack of agreement as <95% of the points lie within the 95% limits of agreement.

Validity

The Bland-Altman plots were used to test the validity of the PAR Index score obtained from digital and plaster models.

When the difference between total PAR Index score from digital and plaster study models were analyzed, 28 out of 30 measurements were within two standard deviations from the mean difference.

Measurements of the individual components that make up the PAR Index score were also analyzed. Four out of the six components that make up the PAR Index score (maxillary anterior region, mandibular anterior region, buccal segment, and midline) had all points within 95% limits of agreement. For overjet and overbite, 28 out of 30 measurements were within 95% limits of agreement. The mean difference and standard deviations are listed in Table 4.

Table 1: Peer Assessment Rating Index score

| Subject	Measurement at T1		Measurement at T2	
	Physical models	Digital models	Physical models	Digital models
1	28	30	30	30
2	46	46	47	46
3	30	30	29	30
4	25	26	26	26
5	50	51	50	49
6	50	49	49	51
7	46	45	44	46
8	64	60	58	57
9	39	39	40	39
10	42	42	41	41
11	38	39	40	37
12	40	41	39	40
13	31	32	31	35
14	45	46	46	47
15	46	47	48	47
16	29	28	28	28
17	31	31	30	31
18	32	33	32	33
19	43	38	38	39
20	21	19	20	19
21	24	24	24	25
22	40	42	45	41
23	40	41	41	42
24	44	44	42	43
25	36	36	36	37
26	48	47	46	46
27	28	30	30	31
28	40	43	40	41
29	31	31	30	31
30	34	35	36	35
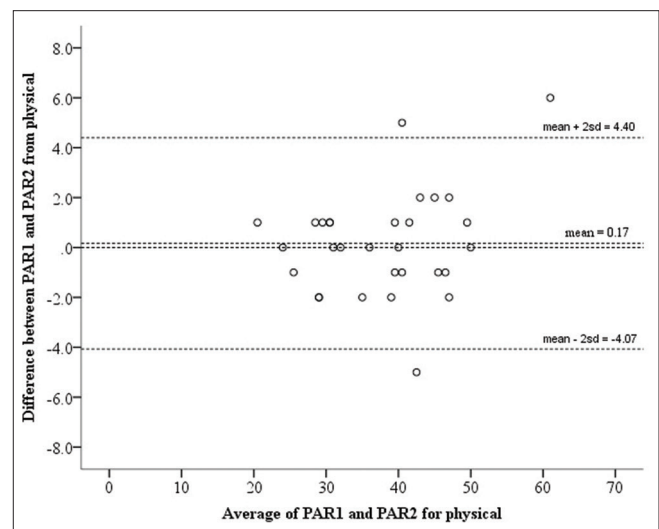
Figure 4: Difference in total weighted peer assessment rating index score between first and second scoring of physical models

Table 2: Data for raw and weighted overjet score

Subject	Raw score (plaster model)	Raw score (digital model)	Difference in raw score	Weighted score (plaster model)	Weighted score (digital model)	Difference in weighted score
1	2	2	0	12	12	0
2	4	4	0	24	24	0
3	1	1	0	6	6	0
4	2	2	0	12	12	0
5	3	3	0	18	18	0
6	3	3	0	18	18	0
7	4	4	0	24	24	0
8	5	4	1	30	24	6
9	3	3	0	18	18	0
10	3	3	0	18	18	0
11	3	3	0	18	18	0
12	3	3	0	18	18	0
13	2	2	0	12	12	0
14	4	4	0	24	24	0
15	4	4	0	24	24	0
16	2	2	0	12	12	0
17	1	1	0	6	6	0
18	3	3	0	18	18	0
19	4	3	1	24	18	6
20	2	2	0	12	12	0
21	2	2	0	12	12	0
22	3	3	0	18	18	0
23	3	3	0	18	18	0
24	4	4	0	24	24	0
25	3	3	0	18	18	0
26	3	3	0	18	18	0
27	2	2	0	12	12	0
28	3	3	0	18	18	0
29	3	3	0	18	18	0
30	3	3	0	18	18	0

Table 3: Results for intra-examiner reliability

	Points within 95% limits of agreement	+2SD	−2SD	Mean
Total PAR Index score (plaster models)	27/30	4.40	−4.07	0.17
Total PAR Index score (digital models)	28/30	2.58	−2.45	0.07

PAR – Peer Assessment Rating; SD – Standard deviation

Table 4: Results for validity

Difference in	Points within 95% limits of agreement	+2SD	−2SD	Mean
Total PAR Index score	28/30*	3.18	−3.44	−0.13
Maxillary anterior	30/30	1.34	−2.41	−0.53
Mandibular anterior	30/30	1.39	−1.39	0
Buccal segment	30/30	1.33	−1.59	−0.13
Overjet	28/30*	3.44	−2.64	0.40
Overbite	28/30*	2.22	−1.95	0.13
Midline	30/30	0	0	0

*<95% of the points lie within the 95% limits of agreement.
PAR – Peer Assessment Rating; SD – Standard deviation

For total PAR Index score, overjet and overbite, there is a lack of agreement in validity as <95% of the points lie within the 95% limits of agreement.

Discussion

Reliability

Richmond *et al*.[7] reported excellent intraexaminer reliability in the measurement of PAR Index score using plaster models. They reported that the maximum difference between the means was 1.53 PAR Index points. Similarly, DeGuzman *et al*.[14] also reported a high intraexaminer agreement in the measurement of PAR Index score on plaster models.

Mayers *et al*.[16] reported excellent intraexaminer reliability in the measurement of PAR Index score for both plaster models and digital models. Similarly, Stevens *et al*.[17] did not detect clinically significant differences in reliability between plaster and digital models in PAR Index measurements.

In our investigation of the intraexaminer reliability of PAR Index score measured from plaster and digital models, for both plaster and digital models, agreement was not obtained as <95% of the points lie within the 95% limits of agreement. For plaster models, three out of thirty samples had the difference between T1 and T2 measurements lie out of the 95% limits of agreement. The mean difference was 0.17 (ranged from 0 to 6) PAR Index points. For digital models, two out of 30 samples had the difference between the first and second reading lie out of the 95% limits of agreement. The mean difference was 0.07 (ranged from 0 to 3) PAR Index points.

The values reported in this study fall within the range published by Stevens *et al.*[17] They reported a mean difference of 2.69 (range from 0.67 to 6.67) UK weighted PAR Index points for plaster models, and a mean difference of 4.56 (range from 0.67 to 14.67) UK weighted PAR Index points for digital models.

Brown and Richmond[18] carried out a calibration study comparing the scores of trainees with the standard score for PAR and Index of Complexity, Outcome and Need (ICON). They reported that the recommended level of acceptable interexaminer agreement for PAR Index score is no more than ± 12. This level is set at perceived level of clinical significance. There have not been published reports on the

recommended level of acceptable intraexaminer agreement of the PAR Index score. Therefore, even though the Bland-Altman Plots show lack of agreement in the measurements, using the recommended acceptable interexaminer agreement of ±12[18] as a guide, the intraexaminer reliability of PAR Index score on both plaster and digital models in this study can be considered clinically acceptable. As Richmond *et al.*[19] stated, absolute agreement cannot be expected, even for experienced users of a clinical index.

Validity

The results of the study showed that there was an agreement in all the measurements between plaster and digital models except for PAR Index score and two of its component scores, overbite and overjet.

Overbite

The difference in the overbite component score of PAR Index, between plaster and digital models, was compared. In this study, two out of 30 samples had differences which lie out of the 95% limits of agreement. The mean difference was 0.13, and the 95% limits of agreement were -1.95 to 2.22.

For these two samples that were not in agreement, the difference in overbite score between the plaster and digital models was two. In the calculation of the PAR index, the overbite component is given a weightage of two.[7] As both the difference in unweighted scores of one and the weighted scores of two are small, the difference in overbite score between plaster and digital models can be considered clinically insignificant.

Published studies have reported clinically insignificant difference in the measurement of overbite between digital and plaster models. Quimby *et al.*[20] reported statistically and clinically insignificant differences in overbite measurement between digital and plaster models. Other studies have reported statistically significant difference in mean overbite measurements of 0.30 mm,[17] 0.49 mm,[2] and 0.21 mm.[21] However, all the authors found the difference clinically insignificant.

In this study, overbite is measured with reference to the amount of coverage of the lower incisors. This is the method of scoring overbite as suggested by Richmond *et al.*[7] To reduce the reported inaccuracies in overbite
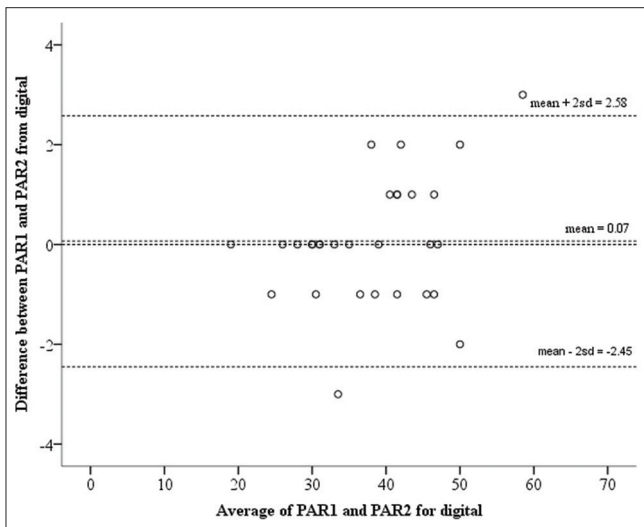

Figure 5: Difference in total weighted peer assessment rating index score between first and second scoring of digital models
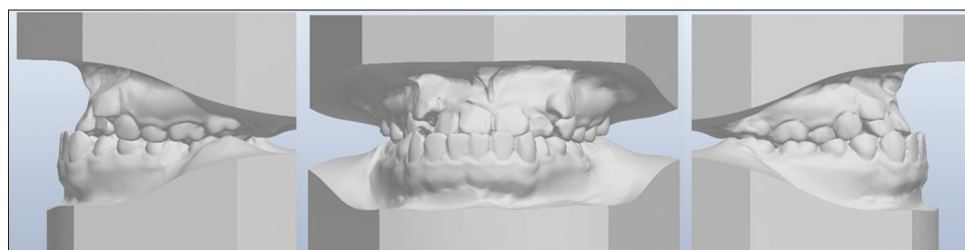

Figure 6: Digital model in standard preset "In-occlusion" buccal and anterior views

measurements that can occur when the digital model is viewed at nonstandardized angles,[2] overbite for digital models is measured with the digital models in the standard preset "in occlusion" anterior view [Figure 6]. This is on the assumption that the standard preset "in occlusion" view sets the occlusal plane of the models parallel to the horizontal. If the occlusal plane of the models is not parallel to the horizontal, a systematic error will be introduced into the vertical and horizontal measurements. This will affect the validity of the digital measurements.

Overjet

The overjet scores of the 30 sets of models are listed in Table 2. The difference in the overjet component score of PAR Index, between plaster and digital models, was compared. In this study, two out of 30 samples had differences which lie out of the 95% limits of agreement. The mean difference was 0.4, and the 95% limits of agreement were -2.64 to 3.44.

When calculating the PAR Index score, the overjet component is given a weightage of six.[7] When the data of the samples that were not in agreement was examined, the difference in overjet score between the plaster and digital models was six. Thus, the actual difference in the unweighted overjet score was one. A one point difference in unweighted score contributed a six-point difference in the weighted overjet component score and accounts for the reason the two samples lie outside the 95% limits of agreement.

The PAR Index ruler was designed for the measurement of the PAR Index score on plaster models. According to Richmond *et al.*,[7] the recording zone for overjet is from the left to right lateral incisors. The most prominent aspect of any one incisor is recorded. When recording the overjet, the PAR Index ruler is held parallel to the occlusal plane and radial to the curvature of the arch [Figure 7].

However, we were unable to apply the same method to measure overjet in digital models as there is no digital equivalent of the PAR Index ruler. The Orthoanalyzer™ program is unable to measure overjet with reference to the curvature of the lower arch form. Overjet was measured by clicking one point on the incisal edge of the upper incisor and one point on the labial surface of the lower incisor. As such, when there is imbrication of the lower incisors [Figure 8], the measured overjet can vary significantly depending on which tooth was selected.

Mayers *et al.*[16] and Stevens *et al.*[17] compared the PAR Index score between plaster and digital models. Both authors did not report any clinically significant difference in the overjet measurements between plaster and digital models. However, the two authors had their own method of determining the overjet component of digital models.

Published studies have reported statistically insignificant difference in the measurement of overjet between digital

Figure 7: Measurement of overjet on plaster model
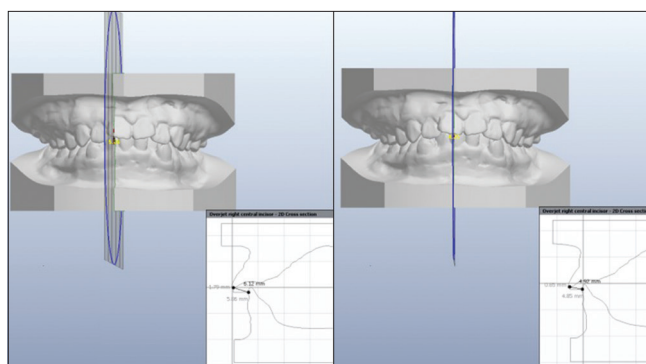

Figure 8: Difference in overjet measurement in presence of lower anterior crowding

and plaster models.[2,17,20] Watanabe-Kanno *et al.*[21] reported a statistically significant but clinically insignificant difference in the mean measurement of overjet at 0.31 mm.

In the measurement of overjet in this study, a difference of 0.31 mm can translate to a difference of one in the raw overjet score, and a difference of six in the weighted overjet score, when the overjet is near the limits of each score range, for example, 3,5,7, and 9 mm. For example, when measuring a plaster model, if the overjet falls on the 5 mm line of the PAR Index ruler, the lower score of one is recorded. For the corresponding digital model, the digital calculation of overjet could be 5.2 mm. This 0.2 mm of difference in overjet between plaster and digital model is clinically insignificant. However, the digital model will score two for overjet, and the resultant difference in overjet component score will be six PAR Index Scores.

Peer Assessment Rating index score

Published studies comparing PAR index score between plaster and digital models reported no clinically significant differences between overall PAR index scores of plaster and digital models.[16,17]

In our investigation of the validity of the PAR index score measured between plaster and digital models, two out of thirty samples had differences which lie out of the 95%

limits of agreement. As <95% of the points lie within the 95% limits of agreement, the Bland-Altman Plot showed that agreement was not attained.

The mean difference was 0.13, and the 95% limits of agreement were -3.44 to 3.18.

For the two samples that were not in agreement, the difference in PAR index score between the plaster and digital models was four and five. When examined, these two samples correspond to the two samples that were out of the 95% limits of agreement in the overjet component. The difference in the PAR index score was largely due to the difference in the overjet component. The large weightage of six points assigned to overjet magnified the difference in the PAR index score between plaster and digital models.

Similarly, even though the Bland-Altman Plots show a lack of agreement in the measurements, using the recommended acceptable interexaminer agreement of ±12[18] as a guide, the validity of PAR index score on digital models compared to plaster models in this study can be considered clinically acceptable.

Statistical analysis

The comparison of the repeatability of each method is relevant to method comparison because the repeatabilities of two methods of measurement limit the amount of agreement that is possible.[22] Lack of agreement in unreplicated studies may suggest that the new method cannot be used, but the reason for this could be due to poor repeatability of the standard method.

If the lack of repeatability and lack of agreement between two methods are similar, the reason for lack of agreement will be due to the lack of repeatability. If the limits of agreement are considerably wider than the repeatability, it suggests that there must be other factors lowering the agreement between methods.[22]

In this study, the 95% limits of agreement for validity (-3.44 to 3.18) are similar to that of repeatability (-4.07 to 4.40 for plaster models and -2.45 to 2.58 for digital models) [Table 5]. Thus, the lack of agreement for validity in the total PAR index score could be due to the lack of agreement for reliability in measurements on both plaster and digital models.

Methodology

In this study, the same plaster model is used for scanning into digital model and for direct measurements. This eliminated the possibility of differences due to distortion of alginate impression.

During sample selection, a visual comparison of the occlusion of each set of plaster model and their corresponding digital model is carried out. If the occlusion of the digital model does not match that of the plaster model, the sample is excluded from the study. Therefore, any difference between plaster and digital models in the study is unlikely to be attributed to inaccuracies in the computerized articulation of the digital models.

When recording the left and right buccal occlusions, care was taken to view the digital models at the standard preset buccal view [Figure 6]. This is because arch forms are curved, and the buccal relationship can appear different when the angle at which the models are viewed changes.

When the digital models are measured, the images are freely rotated and magnified on screen when necessary. However, an enlarged view might make it more difficult for the examiner to consistently and accurately identify the various points to be used for measurements.

For example, contact point displacements are measured by selecting the mesial and distal contact points of adjacent teeth. When the digital model is magnified, the contact point widens to become a contact area. Slightly shifting the position of the chosen points within the contact area will not change the measurement when a PAR ruler is used. However, 3Shape OrthoAnalyzer™ software measures up to an accuracy of 0.01 mm. The small difference in digital measurement can translate to a difference in scoring of the PAR index when the value is near the limits of each score range. The difference in raw PAR score, multiplied by the weightage applied, can magnify the clinically insignificant difference between plaster and digital model measurement.

The examiner's judgment of the exact location of a point may vary at random as it is difficult to select the same points for the measurement on the model each time.[4] Errors in landmark identification can be reduced if markings were placed on the models.[23,24] However, this was not carried out in this study as we were interested in the clinical applicability of digital models in measuring the PAR index score. As some degrees of error are associated with measured data in all experiments, in this study, we expected some random errors.

Table 5: Results for validity and intraexaminer reliability

Intraexaminer reliability	Points within 95% limits of agreement	+2SD	−2SD	Mean
Total PAR Index score (plaster models)	27/30	4.40	−4.07	0.17
Total PAR Index score (digital models)	28/30	2.58	−2.45	0.07
Validity	**Points within 95% limits of agreement**	**+2SD**	**−2SD**	**Mean**
Total PAR Index score	28/30	3.18	−3.44	−0.13

PAR – Peer Assessment Rating; SD – Standard deviation

The sample in this study can be considered a reasonable representation of the range of malocclusions seen in this centre. However, certain traits were not observed or measured. Deciduous teeth and impacted teeth were not included in the sample, and teeth with lateral open bites were not observed.

Future directions

- A digital equivalent of the PAR ruler can be developed for the measurement of the PAR index on digital models
- A PAR calibration course using digital models, which is currently unavailable, can be conducted to standardize the scoring of overjet and overbite on digital models.

Conclusions

1. There was acceptable intraexaminer reliability in the measurement of PAR index score for both plaster and digital models.
2. Agreement was attained between plaster and digital model measurements in four out of the six components that make up the PAR index score. There was a lack of agreement in weighted PAR index score, as well as two of its components: overjet and overbite. However, the differences were clinically acceptable.
3. Digital models are a clinically acceptable alternative to plaster models in the measurement of the PAR index.
4. Improvement in software design and usage are necessary to attain greater agreement in the measurement of the overjet and overbite components of the PAR index score between plaster and digital models.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. McGuinness NJ, Stephens CD. Storage of orthodontic study models in hospital units in the U.K. Br J Orthod 1992;19:227-32.
2. 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.
3. Keating AP, Knox J, Bibb R, Zhurov AI. A comparison of plaster, digital and reconstructed study model accuracy. J Orthod 2008;35:191-201.
4. Abizadeh N, Moles DR, O'Neill J, Noar JH. Digital versus plaster study models: How accurate and reproducible are they? J Orthod 2012;39:151-9.
5. Keim RG, Gottlieb EL, Nelson AH, Vogels DS 3rd. 2011 JCO orthodontic practice study. Part 1 Trends. J Clin Orthod 2011;45:535-44.
6. Shastry S, Park JH. Evaluation of the use of digital study models in postgraduate orthodontic programs in the United States and Canada. Angle Orthod 2014;84:62-7.
7. Richmond S, Shaw WC, O'Brien KD, Buchanan IB, Jones R, Stephens CD, *et al.* The development of the PAR index (Peer assessment rating): Reliability and validity. Eur J Orthod 1992;14:125-39.
8. Richmond S, Shaw WC, Roberts CT, Andrews M. The PAR index (Peer assessment rating): Methods to determine outcome of orthodontic treatment in terms of improvement and standards. Eur J Orthod 1992;14:180-7.
9. Birkeland K, Furevik J, Bøe OE, Wisth PJ. Evaluation of treatment and post-treatment changes by the PAR index. Eur J Orthod 1997;19:279-88.
10. O'Brien KD, Shaw WC, Roberts CT. The use of occlusal indices in assessing the provision of orthodontic treatment by the hospital orthodontic service of England and wales. Br J Orthod 1993;20:25-35.
11. Kerr WJ, Buchanan IB, McColl JH. Use of the PAR index in assessing the effectiveness of removable orthodontic appliances. Br J Orthod 1993;20:351-7.
12. Richmond S, Andrews M. Orthodontic treatment standards in Norway. Eur J Orthod 1993;15:7-15.
13. Pangrazio-Kulbersh V, Kaczynski R, Shunock M. Early treatment outcome assessed by the peer assessment rating index. Am J Orthod Dentofacial Orthop 1999;115:544-50.
14. DeGuzman L, Bahiraei D, Vig KW, Vig PS, Weyant RJ, O'Brien K, *et al.* The validation of the peer assessment rating index for malocclusion severity and treatment difficulty. Am J Orthod Dentofacial Orthop 1995;107:172-6.
15. McGorray SP, Wheeler TT, Keeling SD, Yurkiewicz L, Taylor MG, King GJ, *et al.* Evaluation of orthodontists' perception of treatment need and the peer assessment rating (PAR) index. Angle Orthod 1999;69:325-33.
16. Mayers M, Firestone AR, Rashid R, Vig KW. Comparison of peer assessment rating (PAR) index scores of plaster and computer-based digital models. Am J Orthod Dentofacial Orthop 2005;128:431-4.
17. Stevens DR, Flores-Mir C, Nebbe B, Raboud DW, Heo G, Major PW, *et al.* 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.
18. Brown R, Richmond S. An update on the analysis of agreement for orthodontic indices. Eur J Orthod 2005;27:286-91.
19. Richmond S, Buchanan IB, Burden DJ, O'Brien KD, Andrews M, Roberts CT, *et al.* Calibration of dentists in the use of occlusal indices. Community Dent Oral Epidemiol 1995;23:173-6.
20. Quimby ML, Vig KW, Rashid RG, Firestone AR. The accuracy and reliability of measurements made on computer-based digital models. Angle Orthod 2004;74:298-303.
21. Watanabe-Kanno GA, Abrão J, Miasiro Junior H, Sánchez-Ayala A, Lagravère MO. Reproducibility, reliability and validity of measurements obtained from Cecile3 digital models. Braz Oral Res 2009;23:288-95.
22. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135-60.
23. Asquith J, Gillgrass T, Mossey P. Three-dimensional imaging of orthodontic models: A pilot study. Eur J Orthod 2007;29:517-22.
24. Foong KW, Sandham A, Ong SH, Wong CW, Wang Y, Kassim A, *et al.* Surface laser scanning of the cleft palate deformity – Validation of the method. Ann Acad Med Singapore 1999;28:642-9.