AN ANALYSIS OF INTER-EXAMINER VARIABILITY IN PAR SCORING

by

CRIONA HARTE

BDS (N.U.I), MFDRCSI

A thesis submitted to the faculty of Medicine and Dentistry of the University of Birmingham

for the degree of

MASTER OF PHILOSOPHY

Department of Orthodontics,

School of Dentistry,

St. Chad's Queensway,

Birmingham

B4 6NN

January 2011

ACKNOWLEDGEMENTS

I would like to acknowledge the supervision of Dr A. Dhopatkar in the preparation of this thesis. I greatly appreciate the support and guidance shown by Miss N. Arkutu throughout this study. I wish to thank Dr A. P. White for his statistical advice and assistance.

I am indebted to the staff in the Dental Laboratory, University Hospital of North Staffordshire for their hard work and attention to detail whilst duplicating and preparing the study models.

I would like to thank Ms I. Muqbil whose 'Audit of PAR scoring in the West Midlands 2008' provided data used to perform a sample size calculation in this study. I would also like to acknowledge Mr J. M. Scholey for suggesting this study.

I am grateful for the generosity shown and time given up by all the participants in this study.

ABSTRACT

Objectives

To investigate if there is a statistically significant difference in PAR scoring between two groups of calibrated dental technicians. Group 1 were a national group, recruited from the British Orthodontic Society national list of PAR scoring technicians. Group 2 comprised of a local group of dental technicians in the West Midlands region. In addition, variability in PAR scoring between calibrated dental technicians was investigated using descriptive statistics.

Subjects and Method

Twenty eight examiners were recruited into two equal groups. Each examiner independently observed and PAR scored four completed cases (eight sets of study models). The PAR scores were examined for normality and homogeneity of variance prior to performing a multivariate analysis of variance to test for differences between the two groups.

Results

There was ***no*** significant difference in PAR scoring between the two groups. The PAR scores were normally distributed with equal variance in each group. There was greater variation in PAR scores for pre-treatment study models when compared to post treatment study models for each case.

Conclusions

The PAR scores provided by the technicians in the West Midlands region were comparable to those provided by the technicians nationally. Individual variation between examiners did exist but this variability did not alter the PAR outcome category in 111 out of 112 PAR improvement results obtained.

CONTENTS

CHAPTER ONE

LITERATURE REVIEW

1.1 Defining an Index

An Occlusal Index is a systematic measuring device used in epidemiology and public health to quantify the relationship of teeth and dental arches (Arruda, 2008).

Indices measure malocclusion using numerical or categorical values and enable quantitative or qualitative assessment of malocclusion. Peer Assessment Rating (PAR) Index is an occlusal index that assesses treatment outcome and is therefore utilised to monitor the quality of service provision. Indices which measure treatment need (e.g. Index of Orthodontic Treatment Need, IOTN) enable allocation of limited resources to patients with greatest need.

The requirements of an occlusal index are similar to the requirements of any dental index as summarised in the World Health Organisation Report (WHO, 1966). Summers added validity during time as the tenth point to the list originally proposed by the WHO (Table 1.1). Summers defined validity during time as 'the index concentrating on, and being duly sensitive to, the basic defects of occlusal disorder, rather than to the symptoms of developmental changes' (Summers, 1972).

The major advantage of PAR over previous indices measuring treatment outcome is that it has been shown to be both a valid and reliable measure of treatment outcome (Richmond *et al.*, 1992).

Table 1.1 Requirements for an Occlusal Index, table reproduced from (Tang and Wei, 1993)

1. Status of the group is expressed by a single number which corresponds to a relative position on a finite scale with definite upper and lower limits; running by progressive graduations from zero, i.e., absence of disease, to the ultimate point, i.e., disease in its terminal stage.
2. The index should be equally sensitive throughout the scale.
3. Index value should correspond closely with the clinical importance of the disease stage it represents.
4. Index value should be amendable to statistical analysis.
5. Reproducible.
6. Requisite equipment and instruments should be practicable in actual field situation.
7. Examination procedure should require a minimum of judgement.
8. The index, should be facile enough to permit the study of a large population without undue cost in time or energy.
9. The index would permit the prompt detection of a shift in group conditions, for better or worse.
10. The index should be valid during time.

1.2 Types of Indices

Indices can be broadly categorised as follows:

- **Diagnostic Classification Indices**

  Angles classification

  British Standards Institute classification of Incisor relationship

  Great Ormond Street London and Oslo Yardstick (Goslon)

- **Epidemiologic Indices**

  Decayed/missing/filled teeth (DMFT) Index

- **Treatment Need (treatment priority) Indices**

  Summer's Occlusal Index

  Index of Orthodontic Treatment Need (IOTN)

  Dental Aesthetic Index (DAI)

  Handicapping Labiolingual Deviation (HLD) Index

- **Treatment Outcome Indices**

  Peer Assessment Rating (PAR) Index

  American Board of Orthodontics Objective Grading System (ABO OGS)

- **Treatment Complexity Indices**

  Index of Complexity, Outcome and Need (ICON)

1.3 Development of PAR

In 1986 the 'Schanschieff Report' highlighted a varied standard of orthodontic care in the general dental services. The validity of orthodontic treatment under the NHS was questioned. The Occlusal Index Committee was appointed in 1987 following the 'Schanschieff Report' to develop indices to measure orthodontic treatment need and outcome. This resulted in the development of IOTN and PAR.

PAR was developed in 1987 by the British Orthodontic Standards Working Party which consisted of 10 experienced orthodontists (Richmond *et al.*, 1992). A series of six meetings were convened in 1987. Over 200 study models were discussed and analysed until agreement was reached about specific features deemed to be important in obtaining an estimate of malocclusion. The range of study models examined included various developmental stages along with pre- and post-treatment stages. Each feature was scored and the component scores were summed to give an overall single summary score representing the degree of malocclusion.

A single summary PAR score is generated to reflect the degree of deviation from normal alignment and occlusion. A score of zero indicates perfect alignment and occlusion whilst higher scores (rarely above 50) indicate increasing levels of irregularity and malocclusion. A score of 10 or less is deemed acceptable alignment and occlusion whilst 5 or less suggests almost-ideal occlusion. Richmond *et al.*, define normal occlusion and alignment as 'all

anatomical contact points being adjacent, with a good intercuspal mesh between upper and lower buccal teeth, non-excessive overjet and overbite' (Richmond *et al.*, 1992).

Before PAR, several other occlusal indices were developed to assess the outcome of treatment and treatment success (Berg, 1979, Eismann, 1974, 1980, Gottlieb, 1975). Summers Occlusal Index is an index of treatment need which has been used to assess the outcome of treatment (Elderton and Clark, 1983, 1984, Pickering and Vig, 1975). Unlike previous indices of treatment outcome, the validity and reliability of PAR has been evaluated extensively and published (Richmond *et al.*, 1992).

1.4 Validation of PAR

The validity of an Index describes its ability to measure accurately what it purports to measure (Carlos, 1970). Validation involves comparing the scores with an acceptable gold standard, which is frequently the subjective consensus opinion of a group of experienced orthodontists (Richmond *et al.*, 1992, Younis *et al.*, 1997)

Richmond *et al.,* (1992) described the PAR validation exercise, which was undertaken to assess the extent to which PAR reflected current British orthodontic opinion. The panel of 74 examiners included 22 consultant orthodontists, 22 specialist orthodontists, 11 community dental officers, 15 general dental practitioners and 2 public health orthodontic administrators. Rating scale measurements were recorded to reflect each examiner's opinion on the degree of change due to treatment of 234 start and finish study models. Sixteen cases were duplicated to allow double determination.

The study models were also independently scored by each examiner using PAR (Richmond *et al.*, 1992, Shaw *et al.*, 1995). The agreement between the panel opinion and the PAR index was high. Multiple regression techniques were used to confirm that agreement was further improved by applying weightings to overjet x6, overbite x2 and centreline x4. The collective opinion of the examining panel could thus be applied to the numerical scores generated by the weighted Index.

The findings of this PAR validation exercise confirmed that validity was further improved by the allocation of weighting to certain individual components of the index to reflect their significance. The component score was multiplied by its respective weighting to give a weighted score. The weighted component scores were then summed to generate an overall total weighted PAR score. The application of weighting factors to an occlusal index was not a new phenomenon. Weighting for occlusal indices was first used in the 'Malocclusion Severity Estimate' by Grainger in 1960-1961. Weightings were also applied to the 'Occlusal Index' (Summers, 1971), the 'Treatment Priority Index' (Grainger, 1967) and the 'Handicapping Malocclusion Assessment Record' (Salzmann, 1968).

The PAR weightings described by Richmond *et al.,* (1992) were deemed to reflect 'current British orthodontic opinion' and be flexible to change to reflect future standards or standards in other countries. PAR weightings have however remained unchanged in the UK since they were first introduced over fifteen years ago. Concerns about the high weighting applied to overjet and the equality of applying the same weightings to all malocclusions have been expressed in the literature (Hamdan and Rock, 1999).

British opinion may not reflect the consensus opinion worldwide and consequently PAR has also been validated by American orthodontists. In the US, PAR is validated to reflect both severity of malocclusion and anticipated treatment difficulty (DeGuzman *et al.*, 1995).

In DeGuzman's study, eleven orthodontists examined 200 sets of study models and scored them for malocclusion severity and perceived treatment difficulty. Weightings were calculated from partial regression coefficients to increase the association between the panel's opinion and the PAR score. The resultant total weighted PAR score represents both the perceived malocclusion severity and the treatment difficulty (Table 1.2).

The component weightings applied in America differ from the original weightings developed in the UK (Table 1.3). A weighting factor of 2 is applied for buccal segment occlusion in the US. The weightings for overjet and midline discrepancy are less in the US with a greater weighting applied to overbite when compared to the UK.

Table 1.2 American weightings for malocclusion severity and treatment difficulty

Component	Severity Weighting	Difficulty Weighting	Combined Weighting
Overjet	5	4	4.5
Overbite	3	3	3
Midline discrepancy	3	4	3.5
Buccal Occlusion	2	2	2
Upper anterior alignment	1	1	1

Table 1.3 Weighting comparison for UK and US PAR

Component	UK	US (combined)
Overjet	6	4.5
Overbite	2	3
Midline discrepancy	4	3.5
Buccal Occlusion	-	2
Upper anterior alignment	-	1

1.5 Reproducibility of PAR

A reliable index should be able to measure consistently on different occasions and also when used by different examiners. Reproducibility examines agreement between the trainee and a standard measure. Inter-examiner reliability describes scoring consistency amongst a group of examiners. Richmond *et al.,* (1992a) showed excellent intra- and inter-examiner reliability with intraclass correlation coefficients of R> 0.95 and R=0.91 respectively.

Birkeland *et al.,* (1997) tested inter-examiner agreement between two examiners using intraclass correlation coefficient. This revealed an intraclass coefficient of 0.96. There was no bias between the two examiners.

A study by Pangrazio-Kulbersh *et al.,* (1999) assessed early treatment outcomes using PAR index on 206 study models examined by 10 examiners. The kappa score for inter-examiner reliability was 0.831 and intra-examiner reliability was 0.877 (Pangrazio-Kulbersh *et al.*, 1999).

1.6 Components of PAR Index

The PAR Index consists of 11 components with weightings applied to overjet, overbite and centreline (Richmond *et al.*, 1992).

- Upper right segment x 1
- Upper anterior segment x 1
- Upper left segment x 1
- Lower right segment x 1
- Lower anterior segment x 1
- Lower left segment x 1
- Right buccal occlusion (antero-posterior, vertical and transverse) x 1
- Overjet x 6
- Overbite x 2
- Centreline x 4
- Left buccal occlusion (antero-posterior, vertical and travsverse) x 1

The PAR ruler was developed to enable rapid assessment of a set of study models. Each component of the Index is marked on the ruler to allow rapid assessment and remind the examiner of each component.

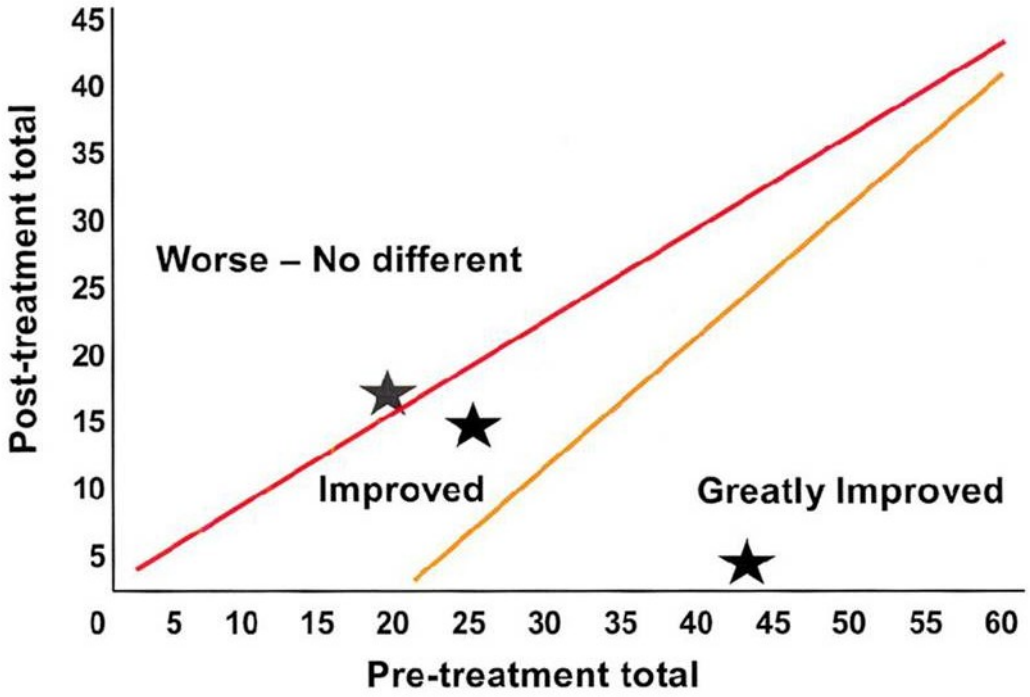
1.7 How PAR measures treatment outcome

Pre- and post-treatment total weighted PAR scores can be compared to reveal a point or percentage score reduction. The degree of PAR improvement can be broadly categorised into the following outcome categories

- *Greatly Improved*: minimum 22 point reduction
- *Improved*: at least 30% reduction
- *Worse or no different*: less than 30% reduction

The treatment outcome category can also be identified using the 'PAR nomogram' (Figure 2). The nomogram was developed using discriminate analysis as part of the validation exercise. The pre-treatment PAR score is plotted on the x-axis against the post-treatment PAR score on the y-axis. The intersect can be determined to identify the PAR outcome category. The degree of improvement is influenced by the pre-treatment PAR score which must be at least 22 points to see a greatly improved result.

Figure 1. PAR nomogram



1.8 Calibration and inter-examiner variability

An Index is defined as a systematic measuring device. For PAR every examiner must be trained and calibrated to a standardized method of measurement.

The aim of PAR calibration is to ensure uniformity and standardisation in assessing the outcome of treatment. Calibration should thus facilitate data comparison between studies, as examiner scoring consistency should be maintained following calibration. Calibration involves attendance at a one day training course. Clinicians, nurses and dental technicians can enrol on a PAR calibration course. It has been shown that non-dental personnel (clerical staff and dental advisers at the Dental Practice Board of England and Wales) could be trained and calibrated to use the PAR Index to a high degree of reliability (Richmond *et al.*, 1993b).

Despite careful training and calibration there is still no guarantee that PAR results will be comparable between calibrated examiners due to examiner bias or human error. The aim of the study is to analyse variability in PAR scoring e.g. due to such examiner bias or human error amongst the sample of calibrated dental technicians recruited. No individual PAR examiner will be considered entirely accurate and variability in PAR scores between examiners due to examiner bias or measurement error is likely.

1.9 Applications of PAR

The Peer Assessment Rating Index is used to monitor and promote standards of orthodontic care.

PAR was used to asses the standard of orthodontic treatment provision in the general dental service in England and Wales (Richmond *et al.*, 1993a, Turbill *et al.*, 1994, Turbill *et al.*, 1996a, b, 1998, 1999). Turbill *et al.,* (1996) showed a mean percentage reduction in weighted PAR score of 47%.

O'Brien *et al.,* (1993) investigated the provision of orthodontic treatment by the hospital orthodontic service of England and Wales. This study showed a mean PAR score reduction of 68% in a sample of 1630 treated cases. Only 57.3% of patients were treated with upper and lower fixed appliances.

An investigation of orthodontic treatment standards in Norway which assessed 200 cases showed a mean PAR score reduction of 78% (Richmond and Andrews, 1993).

Dyken *et al.,* (2001) used the PAR index to compare 54 American Board of Orthodontics accepted cases to 51 cases consecutively treated by orthodontic graduate students. This study showed a mean percentage PAR score reduction of 87.9% for the Board- accepted cases and 81.7% for the graduate student treatment sample. All patients in this study were treated with upper and lower fixed appliances.

Dyken *et al.,* (2001) postulated that different treatment modalities may account for the difference in percentage PAR score reduction seen in the British, Norwegian and American sample.

A greater percentage PAR score reduction has been shown for patients treated with fixed appliances when compared with patients treated with removable appliances (Fox, 1993, Kerr *et al.*, 1993, O'Brien *et al.*, 1993).

Mandall *et al.,* (2010) used PAR to perform the sample size calculation and to assess treatment outcomes in the multi-centre randomised, controlled trial investigating the effectiveness of early class III protraction facemask therapy.

Birkeland *et al.,* (1997) used PAR in a long term study of treatment results assessing changes 5 years post-retention. PAR made it possible to compare treatment standards and long term results post retention for different groups and treatment systems.

Firestone *et al.,* (2002) evaluated PAR as an index of orthodontic treatment need. This study concluded that both the US PAR and the UK PAR scores were excellent predictors of orthodontic treatment need as established by a panel of orthodontists.

A Swedish study comparing PAR to two Swedish indices of treatment need (The modified Indication Index and The modified Index of the Swedish Medical Health Board) found PAR

unsuitable for measuring treatment need since it ‘estimates deviations from an ideal occlusion and the other indices evaluate treatment need’ (Bergstrom and Halling, 1997).

Templeton *et al.,* (2006) concluded that both PAR and the Index of Treatment Complexity, Outcome and Need (ICON) are suitable indices for accessing the clinical outcome for combined orthodontic treatment and orthognathic surgery.

1.10 Limitations of PAR

Hamdan and Rock, (1999) suggested a new weighting system for PAR. The two major limitations highlighted in their paper was the high weight applied to overjet and the application of one weighting system to all malocclusions.

PAR has been weighted to reflect the consensus opinion of British orthodontists but this may not reflect the views of orthodontists in other countries as witnessed by the different weighting applied in the U.S. (DeGuzman *et al.*, 1995).

Occlusal indices do not objectively measure functional, skeletal or psychosocial parameters (McGorray *et al.*, 1999). Changes in facial profile or cephalometric parameters which reflect the skeletal component of malocclusion are not considered (DeGuzman *et al.*, 1995).

PAR is a valid and reliable measure of the dentoocclusal effects of treatment but does not consider treatment difficulty which may influence treatment aims and treatment outcome. A low pre-treatment PAR score does not confer reduced treatment difficulty or treatment time and can preclude a greatly improved result if the pre-treatment score is <22 points.

PAR does not consider iatrogenic effects of treatment such as enamel lesions, marginal bone loss and apical root resorption (Birkeland *et al.*, 1997a, Linge and Linge, 1991).

Dyken *et al.,* (2001) stated that PAR does not evaluate functional occlusion, periodontal health, root resorption, tooth angulations, patient satisfaction or patient compliance.

Treatment duration and patient's satisfaction are not regarded. Fox *et al.,* (1997) suggest 'indices fail to capture the subjective dimensions of health that reflect the patients' values'.

Birkeland *et al.*, (1997) state the PAR Index is not always able to evaluate the outcome of orthodontic treatment effectively. Their investigation evaluating treatment and post-treatment changes showed 3.1 % of cases examined demonstrated no benefit from treatment using PAR. These cases included mild malocclusions, adult cases with a history of multiple early extractions, malocclusions with impacted teeth in the buccal segment and premolar hypodontia. This paper highlights an important issue relating to how the PAR Index scores overjet. An overjet of 3.1-5 mm scores as 6 weighted PAR points, yet an overjet 2-4 mm is often considered as being within normal limits clinically.

Hinman *et al.*, (1995) found PAR to be insensitive when assessing certain aspects of residual treatment need e.g. remaining extraction spaces, rotations and unfavourable incisor inclinations.

PAR is sensitive to malocclusions with increased overjet and weightings mainly favour alignment in the labial segments, with less emphasis on buccal segment relationships. PAR does not consider the position and stability of teeth post-treatment i.e. the potential adverse effects of expansion or proclination during treatment. Malocclusions which score less than 22 points pre-treatment are precluded from a greatly improved result.

1.11 The current status of indices of orthodontic treatment outcome

- PAR

The PAR Index measures treatment success and ultimately clinical performance. It is the most widely accepted index of treatment outcome in orthodontics.

- ICON

The Index of Complexity, Outcome, and Need (ICON) is a multipurpose index developed in 2000 to evaluate orthodontic treatment complexity along with treatment outcome and need (Daniels and Richmond, 2000). It is a weighted index which also includes the aesthetic component of IOTN to assess treatment need. A study by Fox *et al.*, (2002) comparing ICON, PAR and IOTN concluded that ICON may effectively replace IOTN and PAR as a measure of treatment need and outcome respectively. ICON has not as yet however replaced PAR as the recommended measure of orthodontic treatment success in the UK

- ABO-OGS

In 1999 the American Board of Orthodontics instituted an objective grading system (ABO-OGS) for dental casts and radiographs to assess the outcome of treatment as part of the phase III examination (Casko *et al.*, 1998).

- Comparing PAR, ICON and ABO-OGS

Onyeaso and Begole, (2007) investigated the relationship between ICON, Dental Aesthetic Index (DAI), PAR and ABO-OGS. This study showed a significant correlation between DAI and ICON with respect to treatment need. PAR and ABO-OGS also demonstrated significant

correlations with ICON for treatment outcome. The authors concluded that ICON can be used in place of PAR and ABO-OGS for assessing treatment outcome (Onyeaso and Begole, 2007).

- BOS Guidelines

The British Orthodontic Society Clinical Standards Committee guidelines published in July 2009 advocated the use of the PAR Index to assess the treatment outcome of patients. These guidelines, for Primary Care Trusts and local health boards, aim to assess the outcome of patients treated by specialist orthodontists or general dental practitioners (British Orthodontic Society, 2009). For a practitioner to demonstrate high standards, a negligible proportion of their caseload should fall in the 'worse or no different category' (<5%). The mean percentage reduction in PAR score should be high (>70%).

- NHS Orthodontic Contract Requirements

The recent NHS orthodontic contract requires all orthodontists to monitor treatment outcomes for 20 cases, plus 10% of the remainder of their caseload, every year (Contracting, 2009). This document also recommends that Primary Care Trusts should support local contractors by encouraging and facilitating training and calibration in IOTN and PAR.

- Digital Study Models

The transition from plaster study models to digital images will not restrict the application of PAR as a measure of treatment outcome since it has also been shown to be a valid and reliable measure of occlusion on digital study models (Mayers *et al.*, 2005). Stevens *et al.*,

(2006) undertook a study assessing the validity, reliability and reproducibility of plaster versus digital study casts. This revealed no clinically significant difference for the PAR Index and its constituent measurements between plaster casts and digital images.

A Masters thesis presented to Ohio State University by Andrews in 2008 also investigated validity and reliability of PAR scores using digital and plaster study models. This involved ten trained examiners PAR scoring thirty standardised and previously PAR scored study models used in calibration. Both plaster and digital calibration models were scored. This study concluded that validity, intra- and inter examiner reliability of PAR scores on digital and plaster study models is high.

1.12 Summary

The PAR Index has been shown to be a reliable and valid weighted measure of orthodontic treatment success (Richmond *et al.*, 1992). Excellent reliability within and between examiners has been confirmed. To ensure standardisation, PAR relies on examiners completing a PAR calibration course. Despite calibration there is still 'no guarantee that results will be comparable due to differences in experience, personal biases regarding severity or individual aptitude' (Roberts and Richmond, 1997).

Both the 'Quality assurance in NHS primary care orthodontics' and the BOS clinical standards committee guidelines advocate the use of the PAR Index to monitor and assess the quality of orthodontic service provision.

1.13 Aims of study

- To determine if there is a statistically significant difference in PAR scoring between the following two groups of calibrated PAR examiners

  Group 1: 14 dental technicians randomly selected form the BOS national list
  Group 2: 14 dental technicians in the West Midlands region

- To analyse variability in PAR scoring between calibrated dental technicians using descriptive statistics.

The primary objective of this study is to compare PAR scoring in the West Midlands with PAR scoring nationally. This study will also examine variability in PAR scores provided by dental technicians calibrated in the use of the PAR Index.

This is not a test of PAR reliability or reproducibility. The index has been shown in previous studies to be a valid and relaible measure. This study will hopefully provide a snapshot of PAR scoring nationally and this kind of investigation has not been noted previously in the available literature.

CHAPTER TWO

MATERIALS AND METHOD

2.1 Materials

This study, registered with the University of Birmingham, was based at the Orthodontic Department, Birmingham Dental Hospital. The following cost implications and resource requirements were considered.

- Study model duplication, production and packaging costs

Duplication, production and packaging costs were eliminated due to the generosity shown by the staff in the Dental Laboratory, University Hospital of North Staffordshire who kindly agreed to prepare the fourteen duplicate sets of study models required.

- Examiner time and fees for PAR scoring study models

Participants were contacted in writing and invited to participate in this study. One examiner in this study based in a private laboratory requested a fee (£40) for PAR scoring the study models which was paid.

- Study model distribution costs

Study model postage costs were also paid by the School of Dentistry, University of Birmingham. The total cost of distribution was £89.32 (£6.38 x 14). The cost of return postage was kindly covered by the participant laboratories.

2.2 Sample size calculation

The following information was used to perform the sample size calculation

- The minimum difference in scores between the two groups regarded as clinically significant

A review of the literature identified that the 'recommended level of acceptable inter-examiner agreement is no more than +/- 12 points' which is set as perceived level of clinical significance (Brown and Richmond, 2005). Even though this study is not an assessment of inter-examiner agreement, a +/- 12 point difference in the pre-treatment PAR score was chosen to reflect a clinically significant difference between the two groups.

This difference was deemed inappropriate for examining differences in post-treatment PAR scores between the two groups as the range of PAR scores was expected to be much less and seldom greater than 10 points. The following categories of orthodontic treatment results based on final PAR scores were identified in the literature (King *et al.*, 2003)

- Acceptable < or equal 5
- Marginal 5-10
- Poor >10

A minimum difference of +/- 5 points between the two groups was chosen for post-treatment study models.

- An estimate of within-group standard deviation on the PAR score

The following estimates of within-group standard deviation on the PAR score, was obtained from the 2008 West Midlands Audit of PAR scoring data

Pre treatment PAR scores

	Pre1	Pre2	Pre3	Pre4
Examiner 1	38	37	57	35
Examiner 2	38	42	63	38
Examiner 3	37	46	48	32
Examiner 4	37	43	64	41
Examiner 5	37	43	64	41
Examiner 6	35	40	54	41
Examiner 7	38	42	51	35
SD	1.069045	2.794553	6.575568	3.644957

Standard Deviation of variable = 7

Post treatment PAR scores

	Post1	Post2	Post3	Post4
Examiner 1	5	4	12	5
Examiner 2	6	2	10	2
Examiner 3	6	11	14	4
Examiner 4	7	9	11	5
Examiner 5	7	9	11	5
Examiner 6	6	4	13	5
Examiner 7	6	4	9	6
SD	0.690066	3.436499	1.718249	1.272418

Standard Deviation of variable = 3

- Sample Size Calculation

A separate sample size calculation was performed for pre- and post-treatment study models. The significance level of 0.05 was adjusted to account for 4 univariate tests, generating a Bonferroni corrected significance level of 0.0125. The Power of this study is set at 90%.

Pre-treatment study models sample size calculation

Significance Level 0.05, Bonferroni corrected significance of 0.0125

Power 90%

Standard deviation of variable = 7

Clinically relevant difference = 12

= 12 examiners per group

Post-treatment study models sample size calculation

Significance Level 0.05, Bonferroni corrected significance of 0.0125

Power 90%

Standard deviation of variable = 3

Clinically relevant difference = 5

= 14 examiners per group

The sample size calculations determined that 12 examiners per group were required to examine the pre-treatment study models and 14 examiners per group for the post- treatment models. To enable comparison in PAR score reduction data 14 examiners were recruited to examine both pre- and post-treatment study models.

2.3 Study model selection criteria

As the 2008 West Midlands PAR scoring Audit study models were available and data from this Audit had been used in this study, it was decided to use the same study models.

The number of study models was therefore restricted to n = 8 sets. This was deemed reasonable in terms of time taken to PAR score them and also acceptable in terms of production and distribution costs. Statistical advice was sought from the University of Birmingham Statistical Advisory Service. Statistically it would have been preferable to include as many study models as possible. However, it was accepted that due to financial, time and practical distribution constraints, a limited number of study models could be included in this self funded post-graduate research study.

A range of malocclusion types and severity were included. Pre- and post-treatment study models were included to enable analysis of variation in PAR improvement scores and outcome categories.

Inclusion criteria

- Good quality study models, trimmed and based correctly
- Restorations which do not alter tooth size or morphology
- No patient identifier labels

Exclusion criteria

- Cases with craniofacial syndromes or cleft palate
- Extensive restorative dental treatment which results in changes to anatomical number, size, shape or morphology of teeth
- Cases with ectopic/impacted teeth: the initial score may be low and, despite worthwhile treatment, the final score may be no better or even worse.

Case 1 (Appendix 1)

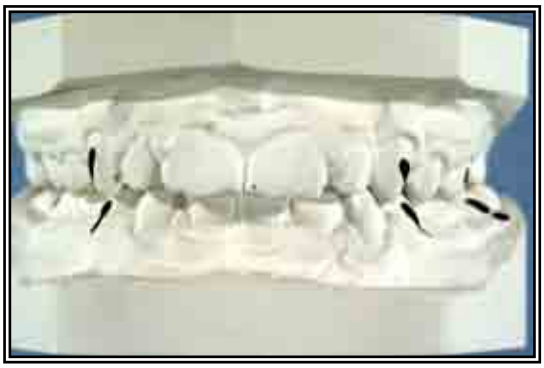
Class III malocclusion with four incisors in crossbite pre-treatment. Treatment included extraction of a premolar unit in each quadrant.

Case 2 (Appendix 2)

Class II division 1 malocclusion with increased overjet treated on a non extraction basis.

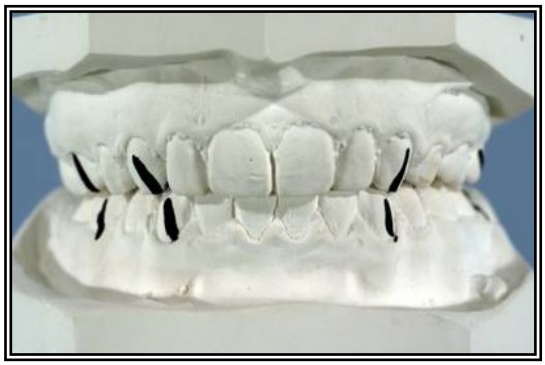
Case 3 (Appendix3)

Class II division 1 malocclusion with severe lower arch crowding, treatment involved extraction of two upper premolars, a lower left premolar and a lingually displaced lower right incisor. A residual overjet was accepted post-treatment.



Case 4 (Appendix 4)

Class II division 2 malocclusion treated with extraction of a premolar unit in each quadrant.



2.4 Ethical issues

The National Research Ethics Service was contacted regarding ethical implications. Following advice the local R&D office was contacted regarding local governance procedures. NHS Ethical Review was deemed not necessary for this study. No ethical implications are expected as the study models are coded and carry no direct patient identifiers. This study registered with South Birmingham PCT was granted Research Management and Governance (RM&G) Permission by Birmingham and the Black Country RM&G Consortium Trusts (Ref no. 1298).

2.5 Examiner Selection

This study was restricted to calibrated dental technicians only. The technicians were invited to participate in writing (Appendix 6) and recruited into two groups, National and West Midlands (Table 2.1).

An up to date list of national calibrated PAR examiners was obtained from the British Orthodontic Society website. This revealed 31 examiners eligible for inclusion in the national group (West Midlands PAR examiners were excluded from the BOS list). The University of Birmingham Statistical Advisory Service performed a computerised equiprobable random selection of 14 examiners from the list. One examiner in this group declined to participate and was thus replaced by another randomly selected examiner.

In the West Midlands three laboratories had more than one participant in this study. Examiners were asked to independently assess and PAR score the study models. The

potential for discussion around a case was unavoidable and considered a limitation of this study. It was however considered an equal risk for all laboratories in this study unless the examiner was working single-handedly in the laboratory and then the potential for discussing the case with colleagues would be less.

Table 2.1 Participant Laboratories

GROUP 1- NATIONAL	GROUP 2- WEST MIDLANDS
Lister Hospital, Stevenage	Birmingham Dental Hospital
Kettering General Hospital	Birmingham Dental Hospital
Royal Lancaster Infirmary	Univ. Hospital of North Staffordshire
JJ Thompson, Sheffield	Univ. Hospital of North Staffordshire
Mexborough Montagu Hospital	Univ. Hospital of North Staffordshire
Leeds Dental Institute	Queen's Hospital , Burton-on-Trent
Sheffield Orthodontics Laboratory	South Warwickshire Foundation Trust
Chiltern Pines Laboratory, Watford	Worcester Royal Hospital
Seacroft Hospital, Leeds	New Cross Hospital , Wolverhampton
OCL (Sheffield) Ltd	Kidderminster Hospital
Broomfield Hospital , Chelmsford	Princess Royal Hospital, Telford
Kingston on Thames Hospital	Royal Shrewsbury Hospital
Queen Victoria Hospital, East Grinstead	Royal Shrewsbury Hospital
Pinderfields Hospital ,Wakefield	Good Hope Hospital , Sutton Coldfield
Dumfries & Galloway Royal Infirmary	

2.6 Statistical analysis

PAR scores are based on ordered categorical scales with values ranging from 0 to above 50 (Roberts and Richmond, 1997). Due to the component weightings, PAR scores were considered as being interval scale measurements. The method of statistical analysis is determined by the type of data which varies for different occlusal indices. 'The IOTN index is categorical in nature whereas both PAR and ICON indices generate more continuous data' (Richmond, 2005).

- Test for normality, Kolmogorov-Smirnov test

- Test for homogeneity of variance, Levene's test statistic

- General linear model multivariate analysis of variance (MANOVA) tests were used to test for significant differences in Pre-treatment, Post-treatment and Improvement PAR scores between the two groups

- Descriptive statistics were used to provide a measure of central tendency and degree of spread in PAR scores for each set of study models in this study

- The PAR nomogram was used to identify if variability in scoring affects the treatment outcome category

2.7 West Midlands Audit of PAR scoring 2008

The aim of this audit was to examine variability in total and component PAR scores for four completed cases (8 study models; 4 pre- and 4 post-treatment). This audit showed variations of >12 points between examiners in two out of eight sets of study models examined. These study models were both pre-treatment, Case 2 PAR score range 33-46 and Case 3 PAR score range 48-64.

This audit showed that variation did exist between calibrated dental technicians in the West Midlands. The audit data provided an estimate of within group standard deviation used for sample size calculation in the present study. Since this audit identified variation between examiners in the West Midlands, the primary objective of the present study was to determine if there is a statistically significant difference in PAR scoring between a group of calibrated dental technicians in the West Midlands and an equal group of calibrated dental technicians nationally.

2.8 Procedure

Fourteen sets of study models were duplicated and packaged in the Dental Laboratory, University Hospital of North Staffordshire. The study models were posted together with data collection sheets (Appendix 5) to each examiner in Group 1 simultaneously.

Each examiner provided component and total weighted PAR scores along with information regarding year of calibration / recalibration using the data collection sheets provided.

Once the study models were returned from the examiners in Group 1 they were then distributed to the examiners locally in Group 2. Prior to redistribution additional duplicate sets were prepared to replace some of the study models which were chipped / damaged.

CHAPTER THREE

RESULTS

Section A Table of Results

Table 3.1 Total Weighted PAR Scores

Examiner	Pre 1	Post 1	Pre 2	Post 2	Pre 3	Post 3	Pre 4	Post 4	Group
1	41	6	52	9	69	21	38	4	1
2	37	3	45	3	47	11	36	2	1
3	41	7	31	3	68	10	37	4	1
4	35	6	32	3	55	11	37	5	1
5	39	5	34	2	64	12	38	4	1
6	37	7	41	5	58	11	32	6	1
7	35	5	43	4	58	8	42	4	1
8	35	4	36	3	54	8	38	5	1
9	36	5	36	6	60	7	41	4	1
10	38	5	42	4	57	14	43	5	1
11	36	4	37	7	57	14	45	3	1
12	39	4	43	2	60	8	20	3	1
13	34	5	31	3	51	8	38	5	1
14	38	5	38	7	35	11	43	5	1
15	32	5	35	4	58	11	44	4	2
16	36	4	42	3	41	8	41	4	2
17	34	4	36	3	64	19	39	4	2
18	39	8	40	6	59	15	35	7	2
19	39	8	39	6	59	15	35	7	2
20	38	8	39	5	59	13	35	6	2
21	36	7	40	6	59	15	35	7	2
22	38	7	35	4	58	12	38	4	2
23	55	5	41	2	45	18	35	3	2
24	38	7	39	10	59	15	35	6	2
25	38	6	44	8	68	11	40	4	2
26	36	4	45	2	60	10	42	4	2
27	36	6	45	4	59	10	42	5	2
28	33	4	54	1	43	1	33	4	2
*	16*	3	37	4	52	9	34	5	1

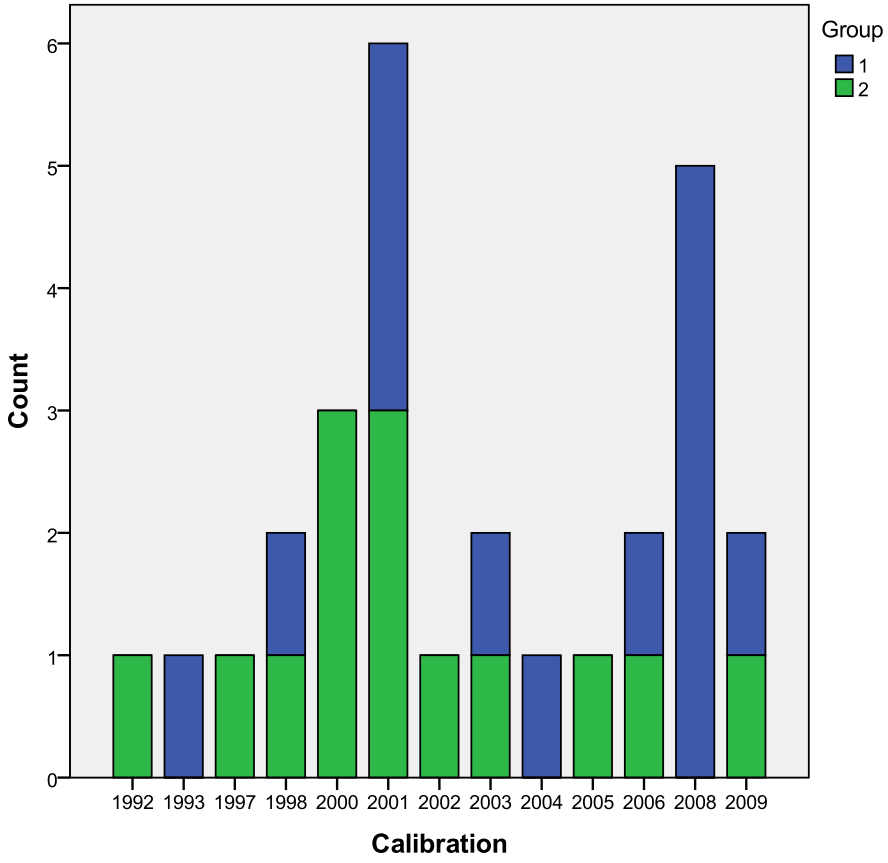
* One anomalous score was obtained for Case I pre treatment. Following statistical advice data from this examiner was not included in the statistical analysis. On the whole data normality assumption was not violated by this decision and effort was made to ensure this by recruiting another examiner to group one.

Table 3.2 PAR Outcome Categories

Examiner	Case 1	Case 2	Case 3	Case 4
1	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
2	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
3	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
4	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
5	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
6	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
7	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
8	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
9	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
10	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
11	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
12	Greatly Improved	Greatly Improved	Greatly Improved	Improved**
13	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
14	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
15	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
16	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
17	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
18	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
19	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
20	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
21	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
22	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
23	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
24	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
25	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
26	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
27	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
28	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved

**low pre-treatment score of 20 meant a 22 point reduction required for greatly improved result was not possible despite a 85% reduction in PAR score.

Figure 3. Year of Calibration



Two examiners provided initial calibration and recalibration dates. In these instances the most recent recalibration date was included (Appendix 10).

Only 32% (9 out of 28 participants) had been calibrated / recalibrated within the last four years and five participants were calibrated more than ten years ago.

Table 3.3 Repeat PAR scores

Examiner	Pre 1	Post 1	Pre 2	Post 2	Pre 3	Post 3	Pre 4	Post 4
A1	38	5	37	4	57	12	35	5
A2	38	7	35	4	58	12	38	4
D1	37	7	43	9	64	11	41	5
D2	38	6	44	8	68	11	40	4
E1	37	7	43	9	64	11	41	5
E2	36	6	45	4	59	10	42	5
F1	38	6	42	4	58	9	35	6
F2	38	8	39	5	59	13	35	6

Repeat measurements were provided by 6 examiners (A-F) on two separate occasions more than two weeks apart. Accounting for random error (2 PAR points) and systematic error (5 PAR points) the repeat measurements were coincident with the original observations.

Section B Statistical Analysis

3.1 Defining the data and data entry

The first step in statistical analysis involved defining the data and identifying the form of each variable. The data collected in this study included the following variables

- Total weighted PAR scores – interval scale measurements
- PAR improvement scores – interval scale measurements
- Outcome category - greatly improved, improved, worse or no different

The PAR score is obtained by adding a set of ordered categorical subcomponents (Roberts and Richmond, 1997). Due to the weightings of the components, PAR scores were considered as being interval scale measurements.

Statistical analysis of the data was performed using the SPSS Statistics 18.0 statistical package under the guidance of the University of Birmingham's Statistical Advisory Service.

3.2 Test for Normality

One-sample Kolmogorov-Smirnov tests confirmed normal distribution ($p>0.05$) of the PAR scores for each set of study models. (Table 3.4 and Table 3.5).

Table 3.4 Kolmogorov-Smirnov Test for Pre-treatment PAR scores

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Pre 1 is normal with mean 37.464 and standard deviation 4.078.	One-Sample Kolmogorov-Smirnov Test	.067	Retain the null hypothesis.
2	The distribution of Pre 2 is normal with mean 39.821 and standard deviation 5.558.	One-Sample Kolmogorov-Smirnov Test	.921	Retain the null hypothesis.
3	The distribution of Pre 3 is normal with mean 56.571 and standard deviation 8.08.	One-Sample Kolmogorov-Smirnov Test	.090	Retain the null hypothesis.
4	The distribution of Pre 4 is normal with mean 37.75 and standard deviation 4.904.	One-Sample Kolmogorov-Smirnov Test	.322	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Table 3.5 Kolmogorov-Smirnov Test for Post-treatment PAR scores

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Post 1 is normal with mean 5.5 and standard deviation 1.427.	One-Sample Kolmogorov-Smirnov Test	.176	Retain the null hypothesis.
2	The distribution of Post 2 is normal with mean 4.464 and standard deviation 2.269.	One-Sample Kolmogorov-Smirnov Test	.274	Retain the null hypothesis.
3	The distribution of Post 3 is normal with mean 11.679 and standard deviation 4.092.	One-Sample Kolmogorov-Smirnov Test	.667	Retain the null hypothesis.
4	The distribution of Post 4 is normal with mean 4.571 and standard deviation 1.26.	One-Sample Kolmogorov-Smirnov Test	.067	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

3.3 Test for Homogeneity of Variance

The variance of data in the two groups should be the same. Levene Statistic was used to assess equality of variances in PAR scores for each study model examined (Tables 3.6 and 3.7). The hypothesis of equal variances in PAR scores in each group is retained for all study models ($p > 0.05$).

Table 3.6 Levene Statistic for Pre-treatment PAR scores

	Levene Statistic	df1	df2	Sig.
Pre 1	.927	1	26	.344
Pre 2	1.095	1	26	.305
Pre 3	.002	1	26	.962
Pre 4	.303	1	26	.587

Table 3.7 Levene Statistic for Post-treatment PAR scores

	Levene Statistic	df1	df2	Sig.
Post 1	3.850	1	26	.061
Post 2	.158	1	26	.695
Post 3	.789	1	26	.383
Post 4	2.833	1	26	.104

3.4 MANOVA comparing Pre-treatment scores between groups

Analysing pre- and post-treatment scores in the same multivariate analysis would violate the homogeneity of variance assumption as variance is greater for pre- than post-treatment scores thus the scores were analyses separately.

A Multivariate test based on a general linear model examining 28 examiners in 2 equal groups with 4 dependent variables Pre 1 – Pre 4 total weighted PAR scores revealed no significant difference in PAR scoring pre-treatment between Group 1 and Group 2. *Hotelling's trace 0.055 not significant (Table 3.8). The Hotelling Trace coefficient is a statistic for a multivariate test of the significance between two groups. It can be thought of as the multivariate analogue of the t-test

Table 3.8 Multivariate test of Pre-treatment PAR scores between Groups

Between-Subjects Factors

		N
Group	1	14
	2	14

Effect		Value	F	Hypothesis df	Error df	Sig.
Group	Pillai's Trace	.053	.319[a]	4.000	23.000	.862
	Wilks' Lambda	.947	.319[a]	4.000	23.000	.862
	Hotelling's Trace*	.055	.319[a]	4.000	23.000	.862
	Roy's Largest Root	.055	.319[a]	4.000	23.000	.862

3.5 MANOVA comparing Post-treatment PAR scores between groups

A Multivariate test based on a general linear model examining 28 examiners in 2 equal groups with 4 dependent variables , Post 1 – Post 4 total weighted PAR scores, revealed no significant difference in PAR scoring post-treatment between Group 1 and Group 2. *Hotelling's trace 0.15 not significant (Table 3.9).

Table 3.9 Multivariate test of Post-treatment PAR scores between Groups

Between-Subjects Factors

		N
Group	1	14
	2	14

Effect		Value	F	Hypothesis df	Error df	Sig.
Group	Pillai's Trace	.131	.864[a]	4.000	23.000	.500
	Wilks' Lambda	.869	.864[a]	4.000	23.000	.500
	Hotelling's Trace*	.150	.864[a]	4.000	23.000	.500
	Roy's Largest Root	.150	.864[a]	4.000	23.000	.500

3.6 MANOVA comparing PAR Improvement scores between groups

A Multivariate test based on a general linear model examining 28 examiners in 2 equal groups with 4 dependent variables, PAR improvement scores for Cases 1 – 4, revealed no significant difference in PAR improvement scores between Group 1 and Group 2. *Hotelling's trace 0.063 not significant (Table 3.10)

Table 3.10 Multivariate test of PAR Improvement scores between Groups

Between-Subjects Factors

		N
Group	1	14
	2	14

Effect		Value	F	Hypothesis df	Error df	Sig.
Group	Pillai's Trace	.059	.363[a]	4.000	23.000	.832
	Wilks' Lambda	.941	.363[a]	4.000	23.000	.832
	Hotelling's Trace*	.063	.363[a]	4.000	23.000	.832
	Roy's Largest Root	.063	.363[a]	4.000	23.000	.832

3.7 Descriptive statistics

In order to condense the data in a meaningful way an average value and a measure of spread of PAR scores for each set of study models was obtained.

Table 3.11 Descriptive Statistics

		Pre 1	Post 1	Pre 2	Post 2	Pre 3	Post 3	Pre 4	Post 4
Mean		37.46	5.50	39.82	4.46	56.57	11.68	37.75	4.57
Median		37.00	5.00	39.50	4.00	58.50	11.00	38.00	4.00
Mode		36[a]	5	36[a]	3	59	11	35	4
Std. Deviation		4.078	1.427	5.558	2.269	8.080	4.092	4.904	1.260
Variance		16.628	2.037	30.893	5.147	65.291	16.745	24.046	1.587
Range		23	5	23	9	34	20	25	5
Minimum		32	3	31	1	35	1	20	2
Maximum		55	8	54	10	69	21	45	7
Percentiles	25	35.25	4.00	36.00	3.00	54.25	8.50	35.00	4.00
	50	37.00	5.00	39.50	4.00	58.50	11.00	38.00	4.00
	75	38.75	7.00	43.00	6.00	60.00	14.75	41.75	5.00

a. Multiple modes exist. The smallest value is shown

- The variance for pre-treatment PAR scores (16.6, 30.9, 65.3, 24) is greater than that for post-treatment PAR scores (2.4, 5.1, 16.7, 1.6).

- Case 3 showed the greatest variance in pre- treatment (65.3; range 34 points) and post-treatment (16.7 ; range 20 points) PAR scores.

3.8 Pre treatment Case 1

Table 3.12 Frequency Table Pre 1

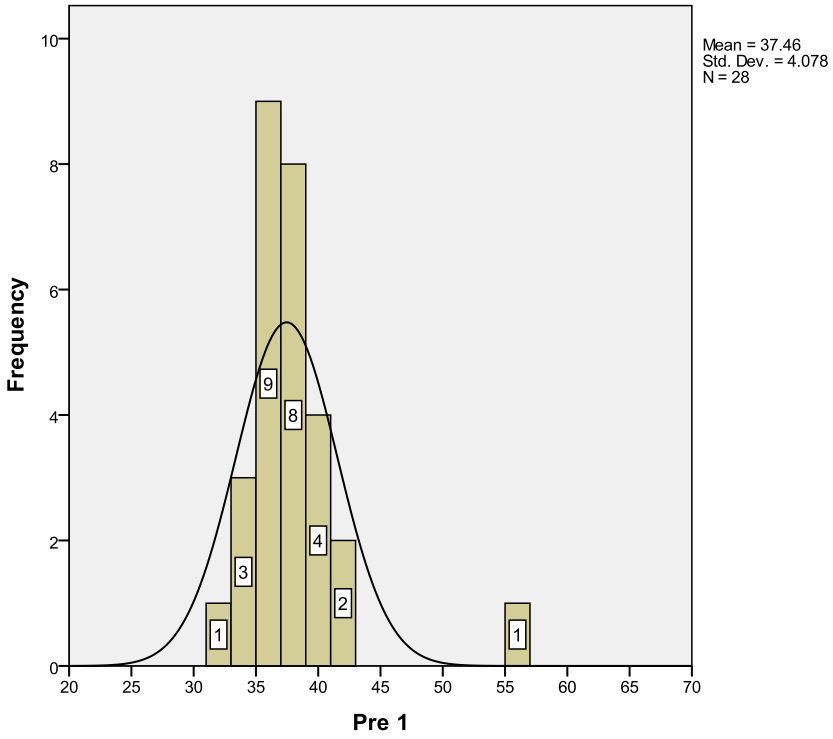
PAR score	Frequency	Percent
32	1	3.6
33	1	3.6
34	2	7.1
35	3	10.7
36	6	21.4
37	2	7.1
38	6	21.4
39	4	14.3
41	2	7.1
55	1	3.6

Table 3.13 Component Scores for Pre 1

		ULS	LLS	Buccal	Overjet	Overbite	Centreline
Mean		4.36	.93	6.39	24.43	1.36	.00
Median		4.00	1.00	6.00	24.00	2.00	.00
Mode		4	1	6[a]	24	0	0
Std. Deviation		.678	.663	1.227	2.268	1.446	.000
Variance		.460	.439	1.507	5.143	2.090	.000
Range		3	2	6	12	4	0
Minimum		3	0	3	24	0	0
Maximum		6	2	9	36	4	0
Percentiles	25	4.00	.25	6.00	24.00	.00	.00
	50	4.00	1.00	6.00	24.00	2.00	.00
	75	5.00	1.00	7.00	24.00	2.00	.00

a. Multiple modes exist. The smallest value is shown

Figure 4. Distribution of Pre 1 PAR scores



The scores obtained for this set of study models were on the whole normally distributed (Mean 37, SD 4, Range 23 points) with one outlier score of 55 obtained. Further analysis of the component PAR scores was undertaken for this case in an attempt to identify the reason for this outlier.

Analysis of the component scores showed the greatest variance in overjet component scores. The examiner who provided a total weighted PAR score of 55 was the only examiner to provide a weighted overjet score of 36.

Table 3.14 Overjet scores

Overjet score	Frequency	Percent
24	27	96.4
36	1	3.6

3.9 Post treatment Case 1

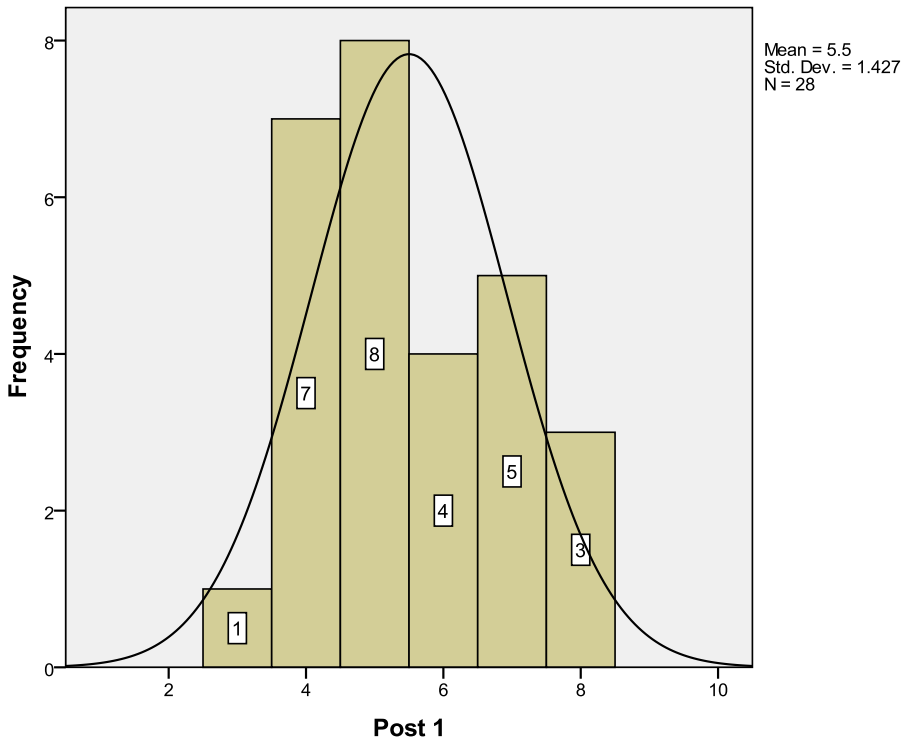
Table 3.15 Frequency Table Post 1

PAR score	Frequency	Percent
3	1	3.6
4	7	25.0
5	8	28.6
6	4	14.3
7	5	17.9
8	3	10.7

Table 3.16 Component Scores for Post 1

		ULS	LLS	Buccal	Overjet	Overbite	Centreline
Mean		.25	.04	5.21	.00	.00	.04
Median		.00	.00	5.00	.00	.00	.00
Mode		0	0	4	0	0	0
Std. Deviation		.585	.189	1.524	.000	.000	.189
Variance		.343	.036	2.323	.000	.000	.036
Range		2	1	5	0	0	1
Minimum		0	0	3	0	0	0
Maximum		2	1	8	0	0	1
Percentiles	25	.00	.00	4.00	.00	.00	.00
	50	.00	.00	5.00	.00	.00	.00
	75	.00	.00	6.75	.00	.00	.00

Figure 5. Distribution of Post 1 PAR scores



The PAR scores for this case were normally distributed with a mean total weighted PAR score of 5, SD 1.4 and range 5 points.

3.10 Pre treatment Case 2

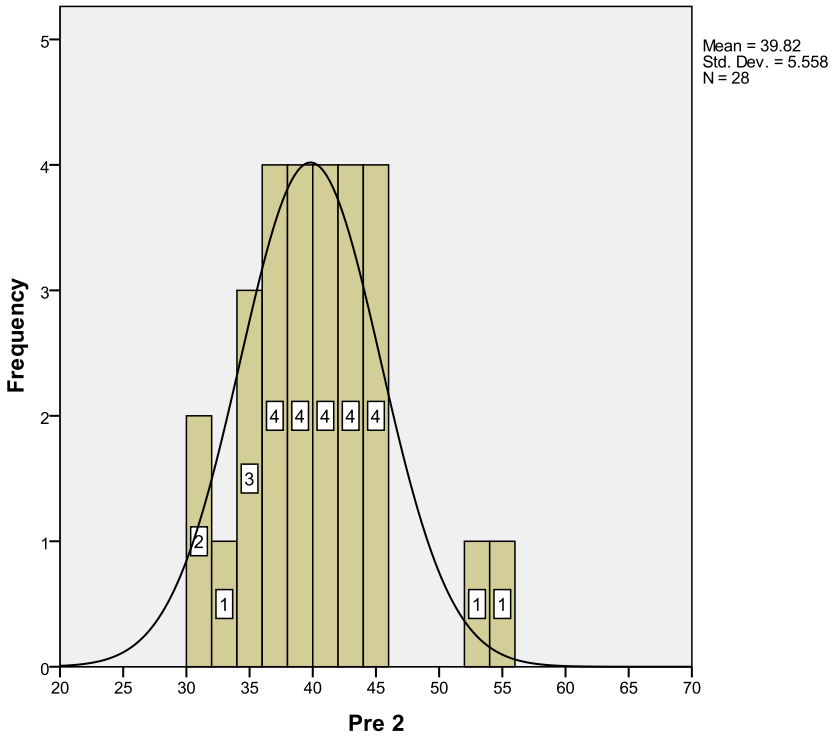
Table 3.17 Frequency Table Pre 2

PAR score	Frequency	Percent
31	2	7.1
32	1	3.6
34	1	3.6
35	2	7.1
36	3	10.7
37	1	3.6
38	1	3.6
39	3	10.7
40	2	7.1
41	2	7.1
42	2	7.1
43	2	7.1
44	1	3.6
45	3	10.7
52	1	3.6
54	1	3.6

Table 3.18 Component Scores for Pre 2

		ULS	LLS	Buccal	Overjet	Overbite	Centreline
Mean		6.18	3.29	2.79	21.64	1.21	4.43
Median		6.00	3.00	2.50	24.00	.00	4.00
Mode		5	3	2	24	0	4
Std. Deviation		1.467	1.049	1.166	3.402	1.572	3.327
Variance		2.152	1.101	1.360	11.571	2.471	11.069
Range		5	5	5	12	6	16
Minimum		4	0	1	18	0	0
Maximum		9	5	6	30	6	16
Percentiles	25	5.00	3.00	2.00	18.00	.00	4.00
	50	6.00	3.00	2.50	24.00	.00	4.00
	75	7.75	4.00	3.75	24.00	2.00	4.00

Figure 6. Distribution of Pre 2 PAR scores



Two outliers were identified (52 and 54). Variation was detected for overjet and centreline scores. The outlier score of 52 was due to an outlying overjet score of 30*. The outlier score of 54 was due to an outlying centreline score of 16**.

Table 3.19 Overjet and centreline scores

Overjet score	Frequency	Percent
18	12	42.9
24	15	53.6
30*	1	3.6

Centreline score	Frequency	Percent
0	5	17.9
4	17	60.7
8	5	17.9
16**	1	3.6

3.11 Post treatment Case 2

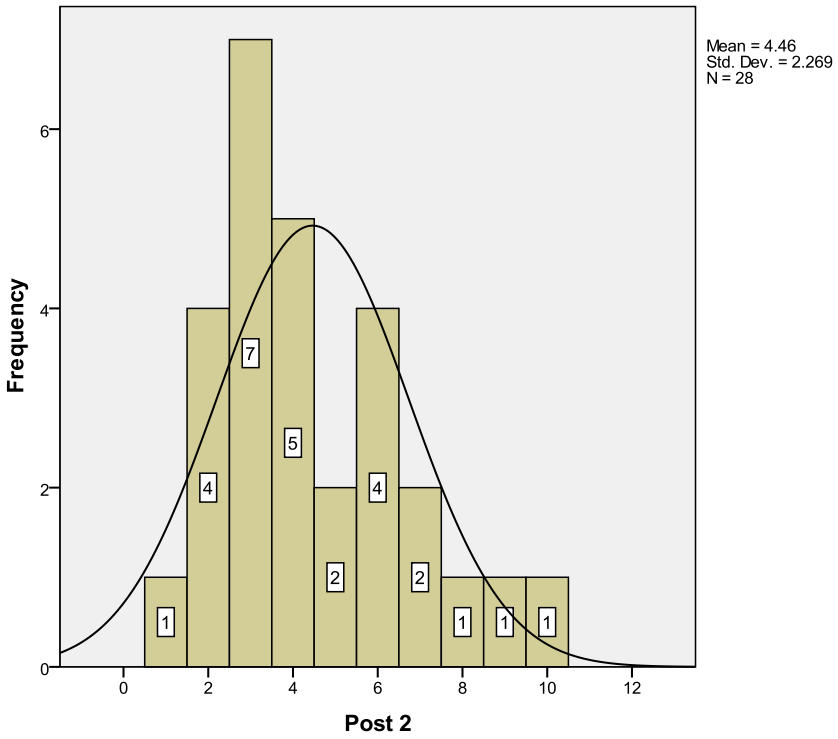
Table 3.20 Frequency Table Post 2

PAR score	Frequency	Percent
1	1	3.6
2	4	14.3
3	7	25.0
4	5	17.9
5	2	7.1
6	4	14.3
7	2	7.1
8	1	3.6
9	1	3.6
10	1	3.6

Table 3.21 Component Scores for Post 2

		ULS	LLS	Buccal	Overjet	Overbite	Centreline
Mean		.79	.29	2.71	.43	.07	.14
Median		1.00	.00	2.50	.00	.00	.00
Mode		1	0	2	0	0	0
Std. Deviation		.686	.810	1.329	1.574	.378	.756
Variance		.471	.656	1.767	2.476	.143	.571
Range		2	4	5	6	2	4
Minimum		0	0	0	0	0	0
Maximum		2	4	5	6	2	4
Percentiles	25	.00	.00	2.00	.00	.00	.00
	50	1.00	.00	2.50	.00	.00	.00
	75	1.00	.00	4.00	.00	.00	.00

Figure 7. Distribution of Post 2 PAR scores



The PAR scores for this case were normally distributed with a mean score of 4, SD 2 and range 9.

3.12 Pre treatment Case 3

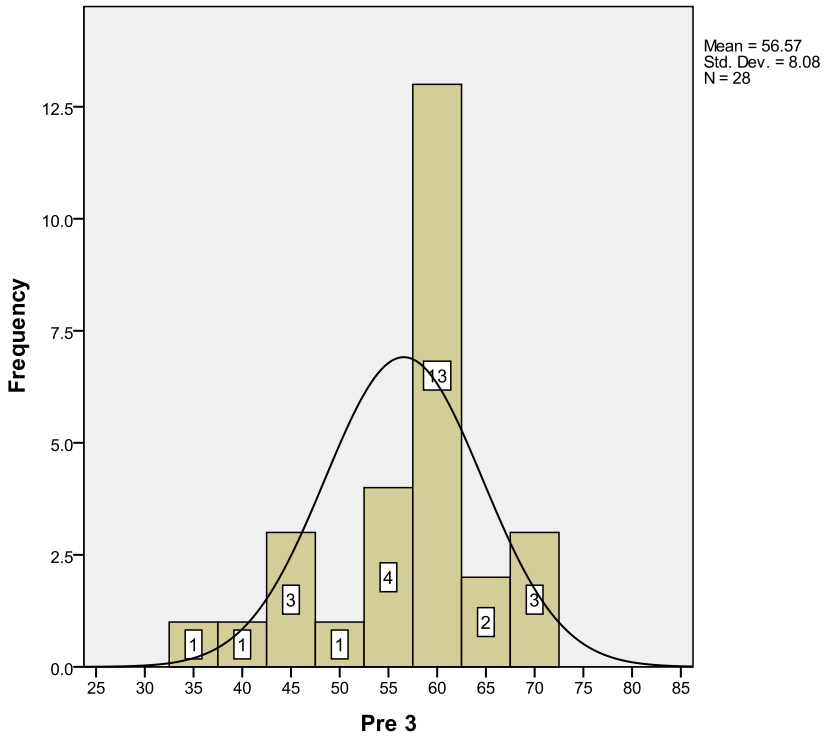
Table 3.22 Frequency Table Pre 3

PAR score	Frequency	Percent
35	1	3.6
41	1	3.6
43	1	3.6
45	1	3.6
47	1	3.6
51	1	3.6
54	1	3.6
55	1	3.6
57	2	7.1
58	4	14.3
59	6	21.4
60	3	10.7
64	2	7.1
68	2	7.1
69	1	3.6

Table 3.23 Component Scores for Pre 3

		ULS	LLS	Buccal	Overjet	Overbite	Centreline
Mean		6.71	10.82	3.75	27.00	3.00	5.29
Median		6.00	10.00	3.00	30.00	2.00	4.00
Mode		6	10	2	30	2	4
Std. Deviation		1.357	1.744	2.137	8.718	1.388	1.902
Variance		1.841	3.041	4.565	76.000	1.926	3.619
Range		6	9	7	30	4	4
Minimum		5	9	0	6	2	4
Maximum		11	18	7	36	6	8
Percentiles	25	6.00	10.00	2.00	30.00	2.00	4.00
	50	6.00	10.00	3.00	30.00	2.00	4.00
	75	7.75	11.00	5.75	30.00	4.00	8.00

Figure 8. Distribution of Pre 3 PAR scores



This case showed a variance of 76 PAR points for the overjet score with a range of weighted overjet values from 6 to 36 PAR points in this Class II division I malocclusion with increased overjet and anterior crossbite pre-treatment.

Table 3.24 Overjet scores

Overjet score	Frequency	Percent
6	2	7.1
12	3	10.7
24	1	3.6
30	18	64.3
36	4	14.3

3.13 Post treatment Case 3

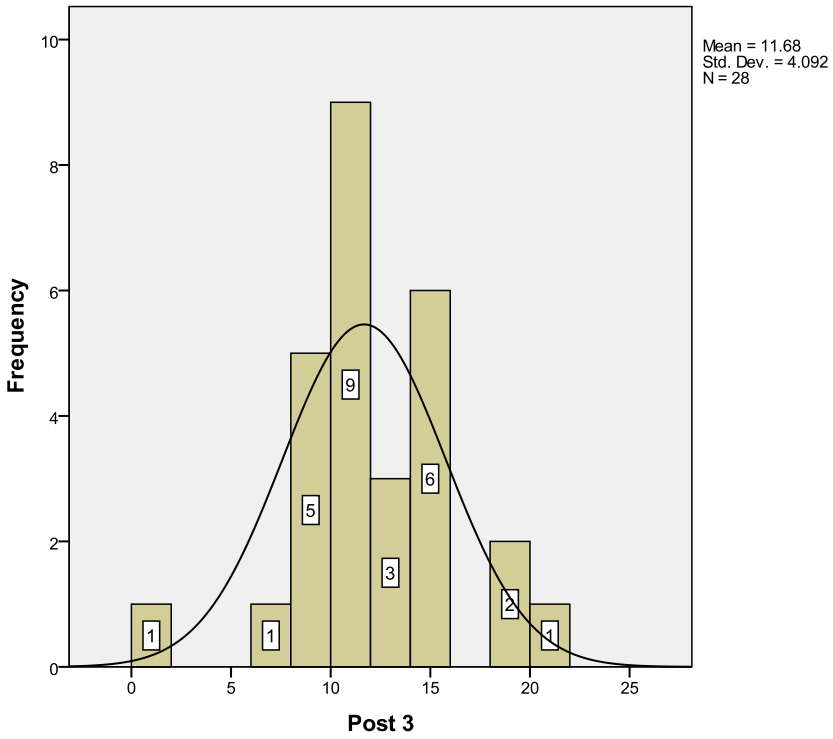
Table 3.25 Frequency Table Post 3

PAR score	Frequency	Percent
1	1	3.6
7	1	3.6
8	5	17.9
10	3	10.7
11	6	21.4
12	2	7.1
13	1	3.6
14	2	7.1
15	4	14.3
18	1	3.6
19	1	3.6
21	1	3.6

Table 3.26 Component scores for Post 3

		ULS	LLS	Buccal	Overjet	Overbite	Centreline
Mean		.14	.46	3.21	6.07	1.29	.57
Median		.00	.00	2.50	6.00	2.00	.00
Mode		0	0	2	6	2	0
Std. Deviation		.448	1.347	2.079	2.142	1.117	1.794
Variance		.201	1.813	4.323	4.587	1.249	3.217
Range		2	7	7	12	4	8
Minimum		0	0	0	0	0	0
Maximum		2	7	7	12	4	8
Percentiles	25	.00	.00	2.00	6.00	.00	.00
	50	.00	.00	2.50	6.00	2.00	.00
	75	.00	.75	4.75	6.00	2.00	.00

Figure 9. Distribution of Post 3 PAR scores



This case showed a mean total weighted PAR score of 11 points, SD 4 and range 20 points. It was not expected to find this degree of variance in PAR scoring a set of post treatment study models. This case was a Class II division 1 malocclusion and treatment included extraction of a lower incisor with a residual overjet accepted post treatment. The high scores suggest a poor treatment outcome for this case. The variance is due mainly to differences in scoring buccal segment occlusion (variance 4), overjet (variance 4) and centreline (variance 3) components.

3.14 Pre treatment Case 4

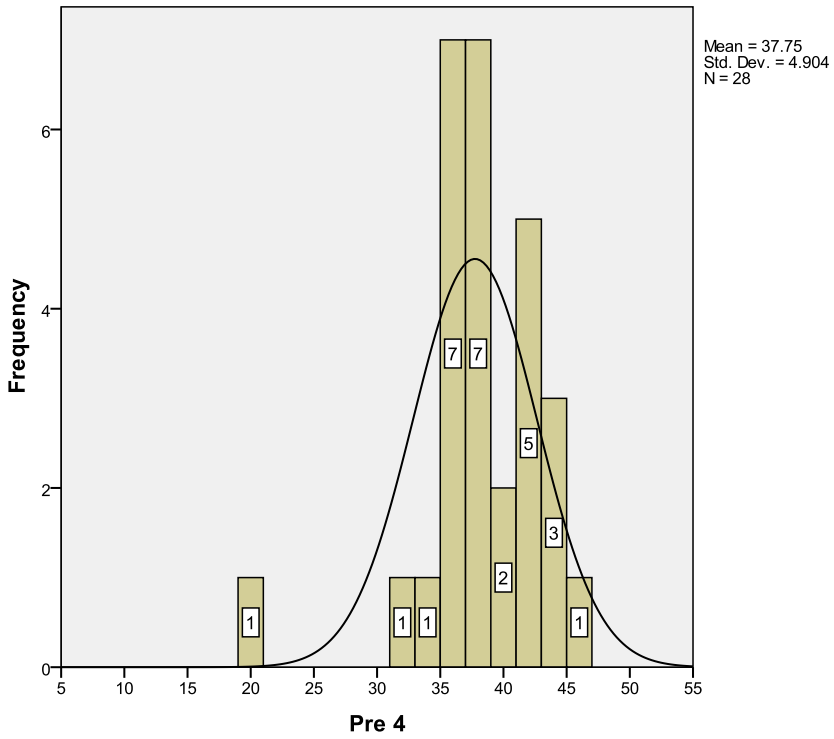
Table 3.27 Frequency Table Pre 4

PAR score	Frequency	Percent
20	1	3.6
32	1	3.6
33	1	3.6
35	6	21.4
36	1	3.6
37	2	7.1
38	5	17.9
39	1	3.6
40	1	3.6
41	2	7.1
42	3	10.7
43	2	7.1
44	1	3.6
45	1	3.6

Table 3.28 Component Scores Pre 4

		ULS	LLS	Buccal	Overjet	Overbite	Centreline
Mean		6.93	4.07	2.54	20.14	3.50	.57
Median		6.50	4.00	2.00	18.00	4.00	.00
Mode		6	4	2	18	2	0
Std. Deviation		1.386	.466	1.261	4.071	2.009	1.425
Variance		1.921	.217	1.591	16.571	4.037	2.032
Range		6	2	4	18	8	4
Minimum		5	3	1	6	0	0
Maximum		11	5	5	24	8	4
Percentiles	25	6.00	4.00	2.00	18.00	2.00	.00
	50	6.50	4.00	2.00	18.00	4.00	.00
	75	8.00	4.00	3.00	24.00	5.50	.00

Figure 10. Distribution of Pre 4 PAR scores



One outlier score of 20 was obtained for this case. The highest variance was seen with overjet component scores and data analysis confirmed that the outlier score of 20 was obtained due to a low overjet score of 6. This was the only examiner to provide this score.

Table 3.29 Overjet scores

Overjet score	Frequency	Percent
6	1	3.6
18	15	53.6
24	12	42.9

3.15 Post treatment Case 4

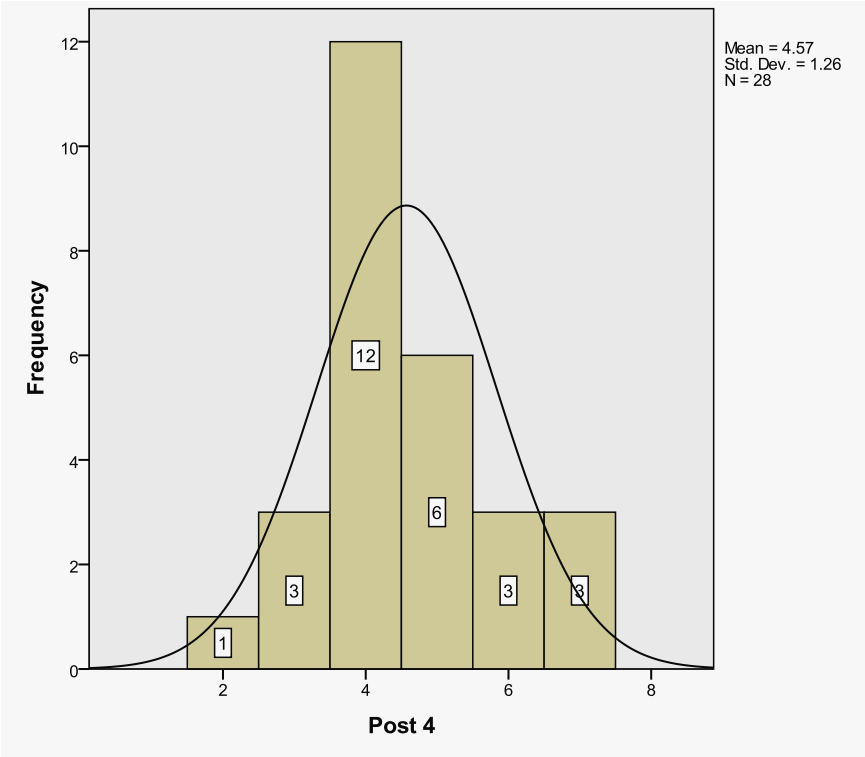
Table 3.30 Frequency Table Post 4

PAR score	Frequency	Percent
2	1	3.6
3	3	10.7
4	12	42.9
5	6	21.4
6	3	10.7
7	3	10.7

Table 3.31 Component Scores Post 4

		ULS	LLS	Buccal	Overjet	Overbite	Centreline
Mean		.46	.04	4.07	.00	.00	.00
Median		.00	.00	4.00	.00	.00	.00
Mode		0	0	4	0	0	0
Std. Deviation		.508	.189	1.412	.000	.000	.000
Variance		.258	.036	1.995	.000	.000	.000
Range		1	1	5	0	0	0
Minimum		0	0	2	0	0	0
Maximum		1	1	7	0	0	0
Percentiles	25	.00	.00	3.00	.00	.00	.00
	50	.00	.00	4.00	.00	.00	.00
	75	1.00	.00	4.75	.00	.00	.00

Figure 11. Distribution of Post 4 PAR scores



The PAR scores for this case were normally distributed with a mean 4, SD 1 and range 5.

3.16 PAR outcome categories

PAR Improvement can be broadly categorised as

- Greatly Improved
- Improved
- Worse or no different

Case 1

100% of examiners (n=28) categorised the PAR outcome as greatly improved.

Case 2

100% of examiners (n=28) categorised the PAR outcome as greatly improved.

Case 3

100% of examiners (n=28) categorised the PAR outcome as greatly improved.

Case 4

27 examiners categorised the PAR outcome as greatly improved.

1 examiner categorised the PAR outcome for this case as improved

Pre- treatment score 20

Post-treatment score 3

85% reduction in PAR score

This examiners low pre-treatment score of 20 meant a 22 point reduction required for greatly improved result was not possible for this case despite a 85% reduction in PAR score.

Twenty eight examiners PAR scored four cases and 111/112 outcomes were categorised as greatly improved which shows excellent consistency in scoring.

CHAPTER FOUR

DISCUSSION

Discussion

The PAR Index is increasingly being used to assess treatment outcome. Despite the development of ICON as an effective replacement measure, PAR is still the most widely accepted measure of treatment success. It plays a very important role in commissioning and monitoring the quality of orthodontic service provision.

PAR relies on calibration to ensure standardisation. The objective of this study was to compare PAR scores obtained in the West Midlands region to those provided nationally and to obtain a 'snap-shot' of the PAR scoring service provided by calibrated dental technicians nationally.

This study shows no significant difference in PAR scoring between the two groups of calibrated dental technicians. The PAR scores provided by the dental technicians in the West Midlands region were comparable to those provided by a random sample of dental technicians nationally randomly selected from the British Orthodontic Society list of calibrated PAR scorers. Local laboratories in the West Midlands region are providing a PAR scoring service which is in line with dental laboratories nationally. All participants in this study were calibrated and the PAR calibration process has effectively ensured standardisation between the two groups in this study.

The present study identified a variation in PAR scoring between individual examiners, as expected. This may be due to bias, measurement error, differences in opinion or

unexplained random variation. This variation did not alter the treatment outcome category in 99% (n = 111/112) of cases examined.

Descriptive statistics identified more variation in pre-treatment scores compared to post-treatment scores and this would also be expected since the range of post-treatment PAR scores would be less.

The PAR scores for each case were normally distributed, however, outliers were identified. Further analysis of the PAR component scores revealed the outliers were largely due to differences in scoring overjet. The weighting factor for overjet is relatively high at 6 and this magnifies the difference between examiners when looking at total weighted PAR scores.

Greatest variation in PAR scoring was seen for Case 3 (Appendix 3). This was a class II division 1 malocclusion with an increased overjet, anterior crossbite and centreline discrepancy. These features of the malocclusion score high when weighting factors are considered and thus it was not surprising to see the greatest variation in total weighted PAR scores for this case. Despite variations in total weighted PAR scores, this case was categorised as greatly improved by all examiners.

It is important that potential for measurement error should be considered in this study. PAR is subjective and identifying measurement error is difficult without first obtaining a 'gold standard' score. No single PAR examiner can be deemed entirely correct and a consensus opinion is frequently necessary to obtain a 'gold standard' PAR score.

A ‘gold standard’ score was not deemed necessary for the study models in this study as this study did not aim to test examiner reliability. The PAR Index has previously been shown to be both a valid and reliable measure and an insufficient number of study models were examined in this study to perform reliability tests statistically. Thus it is not pertinent to draw conclusions with regard to inter- or intra-examiner reliability or agreement from this study. Funding limitations and logistical issues relating to production and distribution of more study models along with time constraints for the participants who willingly agreed to perform the PAR scoring free of charge prevented the inclusion of more study models in this study.

Another consideration is the possible influence of being a participant in a study and the effect, if any, which this may have on PAR scoring. The examiners were aware that the scores they provided would form part of a study investigating variations in PAR scoring. It could be possible that this may influence the way in which they PAR scored the study models in this study i.e. take longer or with more attention to detail or the possibility to discuss the case with colleagues. Thus the procedure may not reflect their everyday practice in PAR scoring. For the purposes of this study these potential limitations were considered and deemed unavoidable.

One anomalous score was however identified which did not reflect the scores obtained from other examiners. This score was for Case 1 pre-treatment (Appendix 1) where one examiner provided an incorrect unweighted reverse overjet score of 1 for four teeth in crossbite. Greater than two teeth in crossbite should automatically result in a reverse overjet score of 4,

as scored by all the other examiners in the study. This was the only occasion where measurement error was identified in this study.

A number of study models were chipped or damaged on return from Group 1. This did not appear to affect the PAR scoring process and it was not reported or documented in the data collection sheets. To avoid this problem the majority of study models for Group 2 were not posted, but delivered by Specialist Registrars working in their respective District General Hospitals in the West Midlands region.

It was expected that more examiners would have undergone recalibration recently. Approximately one third of the examiners had been calibrated or recalibrated in the past four years (Fig. 3). The year of calibration for this sample of dental technicians ranged from 1992 to 2009 with only two technicians undertaking recalibration during this period. Despite this 17 year span during which the participants were calibrated the scores obtained in this study were normally distributed with equal variance in each group. Scoring consistency therefore appears to be maintained over time following calibration and this may be due to frequent application and use of the Index.

PAR is fundamental to the local quality assurance requirement to the new orthodontic contract. The 'Quality assurance in NHS primary care orthodontics' document advises Primary Care Trusts to support local contractors. This includes providing encouragement in training and calibration in the use of PAR and IOTN as part the three tiered approach to

local quality assurance which encompasses mandatory monitoring, peer review and self-regulation.

The findings of this study suggest that the PAR training and calibration process appears to be effective in providing a standardised measure of treatment success and scoring consistency appeared to be maintained over time by the examiners in this study.

Prior to this study there were no data available in the literature regarding variability in PAR scoring between calibrated dental technicians. The present study shows ***no*** significant difference in PAR scoring between two groups of calibrated dental technicians, despite the 17 year span during which the participants had been calibrated.

Suggestions for further research

A copy of the findings of this study will be distributed to each examiner. A questionnaire regarding recalibration will be included. It would be beneficial to obtain the consensus opinion of this sample of calibrated dental technicians on the subject of recalibration since two thirds of this sample had not been recalibration in the past four years.

The findings of this study show that despite calibration individual inter-examiner variation in PAR scoring did occur but the PAR outcome category was unaffected in 99% of cases. At present there are no national guidelines or requirements for examiners providing a PAR scoring service to undertake recalibration. If the PAR outcome category is used to assess treatment success, accepting individual inter-examiner variation in pre- and post-treatment scores, then recalibration may not be essential to maintain scoring consistency over time for calibrated dental technicians.

Further research to compare previously calibrated (> 4 years) to recently recalibrated examiners would be beneficial as it would provide an indication of PAR scoring consistency over time. It would be necessary to use study models with agreed 'gold standard' PAR scores to examine the effect of recalibration on scoring accuracy over time.

CHAPTER FIVE

CONCLUSIONS

Conclusions

1. There was ***no*** significant difference in PAR scoring between the two groups of examiners in this study.

2. 99% (n=111/112) of PAR improvement scores in this study were classified as greatly improved. Variability in PAR scoring generated a different PAR outcome category (Improved) in 1 out of 112 PAR improvement results obtained.

3. There was greater variation in PAR scores for pre-treatment study models when compared to post treatment study models for each case.

4. Variations in overjet component scores were largely responsible for outlier PAR scores.

REFERENCES

References

Arruda, A.O. (2008) Occlusal indexes as judged by subjective opinions. **Am J Orthod Dentofacial Orthop**, 134: (5): 671-675.

Berg, R. (1979) Post-retention analysis of treatment problems and failures in 264 consecutively treated cases. **Eur J Orthod**, 1: (1): 55-68.

Bergstrom, K. and Halling, A. (1997) Comparison of three indices in evaluation of orthodontic treatment outcome. **Acta Odontologica Scandinavica**, 55: (1): 36-43.

Birkeland, K., Boe, O.E. and Wisth, P.J. (1997a) Subjective assessment of dental and psychosocial effects of orthodontic treatment. **J Orofac Orthop**, 58: (1): 44-61.

Birkeland, K., Furevik, J., Boe, O.E., *et al.* (1997b) Evaluation of treatment and post-treatment changes by the PAR Index. **Eur J Orthod**, 19: (3): 279-288.

British Orthodontic Society, C.S.C. (2009) "Guidelines for Primary Care Trusts and Local Health Boards to assess the treatment outcome of patients treated by specialist orthodontists or dentists using the Peer Assessment Rating (PAR) Index ".

Brown, R. and Richmond, S. (2005) An update on the analysis of agreement for orthodontic indices. **Eur J Orthod**, 27: (3): 286-291.

Buchanan, I.B., Shaw, W.C., Richmond, S., *et al.* (1993) A comparison of the reliability and validity of the PAR Index and Summers' Occlusal Index. **Eur J Orthod**, 15: (1): 27-31.

Carlos, J.P. (1970) Evaluation of indices of malocclusion. **Int Dent J**, 20: (4): 606-617.

Casko, J.S., Vaden, J.L., Kokich, V.G., *et al.* (1998) Objective grading system for dental casts and panoramic radiographs. **Am J Orthod Dentofacial Orthop**, 114: (5): 589-599.

Contracting, N.P.C. (2009) "Quality assurance in NHS primary care orthodontics". In Contracting, N.P.C. (Ed.).

Daniels, C. and Richmond, S. (2000) The development of the index of complexity, outcome and need (ICON). **J Orthod**, 27: (2): 149-162.

Deguchi, T., Honjo, T., Fukunaga, T., *et al.* (2005) Clinical assessment of orthodontic outcomes with the peer assessment rating, discrepancy index, objective grading system, and comprehensive clinical assessment. **Am J Orthod Dentofacial Orthop**, 127: (4): 434-443.

DeGuzman, L., Bahiraei, D., Vig, K.W., *et al.* (1995) The validation of the Peer Assessment Rating index for malocclusion severity and treatment difficulty. **Am J Orthod Dentofacial Orthop**, 107: (2): 172-176.

Dyken, R.A., Sadowsky, P.L. and Hurst, D. (2001) Orthodontic outcomes assessment using the peer assessment rating index. **Angle Orthod**, 71: (3): 164-169.

Eismann, D. (1974) A method of evaluating the efficiency of orthodontic treatment. **Trans Eur Orthod Soc**, 223-232.

Eismann, D. (1980) Reliable assessment of morphological changes resulting from orthodontic treatment. **Eur J Orthod**, 2: (1): 19-25.

Elderton, R.J. and Clark, J.D. (1983) Orthodontic treatment in the General Dental Service assessed by the Occlusal Index. **Br J Orthod**, 10: (4): 178-186.

Elderton, R.J. and Clark, J.D. (1984) An investigation of treatment provided in the General Dental Service for patients with Class II division 1 malocclusions. **Br J Orthod**, 11: (1): 2-8.

Firestone, A.R., Beck, F.M., Beglin, F.M., *et al.* (2002) Evaluation of the peer assessment rating (PAR) index as an index of orthodontic treatment need. **Am J Orthod Dentofacial Orthop**, 122: (5): 463-469.

Fox, N.A. (1993) The first 100 cases: a personal audit of orthodontic treatment assessed by the PAR (peer assessment rating) index. **Br Dent J**, 174: (8): 290-297.

Fox, N.A., Daniels, C. and Gilgrass, T. (2002) A comparison of the index of complexity outcome and need (ICON) with the peer assessment rating (PAR) and the index of orthodontic treatment need (IOTN). **Br Dent J**, 193: (4): 225-230.

Fox, N.A., Richmond, S., Wright, J.L., *et al.* (1997) Factors affecting the outcome of orthodontic treatment within the general dental service. **Br J Orthod**, 24: (3): 217-221.

Gottlieb, E.L. (1975) Grading your orthodontic treatment results. **J Clin Orthod**, 9: (3): 155-161.

Grainger, R.M. (1967) Orthodontic treatment priority index. **Vital Health Stat 2**, (25): 1-49.

Hamdan, A.M. and Rock, W.P. (1999) An appraisal of the Peer Assessment Rating (PAR) Index and a suggested new weighting system. **Eur J Orthod**, 21: (2): 181-192.

Hinman, C. (1995) The Dental Practice Board. Orthodontics--the current status. **Br J Orthod**, 22: (3): 287-290.

Kerr, W.J., Buchanan, I.B. and McColl, J.H. (1993) Use of the PAR index in assessing the effectiveness of removable orthodontic appliances. **Br J Orthod**, 20: (4): 351-357.

King, G.J., McGorray, S.P., Wheeler, T.T., *et al.* (2003) Comparison of peer assessment ratings (PAR) from 1-phase and 2-phase treatment protocols for Class II malocclusions. **Am J Orthod Dentofacial Orthop**, 123: (5): 489-496.

Linge, L. and Linge, B.O. (1991) Patient characteristics and treatment variables associated with apical root resorption during orthodontic treatment. **Am J Orthod Dentofacial Orthop**, 99: (1): 35-43.

Mandall, N., DiBiase, A., Littlewood, S., *et al.* (2010) Is early class III protraction facemask treatment effective? A multicentre, randomized, controlled trial: 15-month follow-up. **J Orthod**, 37: (3): 149-161.

Mayers, M., Firestone, A.R., Rashid, R., *et al.* (2005) Comparison of peer assessment rating (PAR) index scores of plaster and computer-based digital models. **Am J Orthod Dentofacial Orthop**, 128: (4): 431-434.

McGorray, S.P., Wheeler, T.T., Keeling, S.D., *et al.* (1999) Evaluation of orthodontists' perception of treatment need and the peer assessment rating (PAR) index. **Angle Orthod**, 69: (4): 325-333.

O'Brien, K.D., Shaw, W.C. and Roberts, C.T. (1993) The use of occlusal indices in assessing the provision of orthodontic treatment by the hospital orthodontic service of England and Wales. **Br J Orthod**, 20: (1): 25-35.

Onyeaso, C.O. and Begole, E.A. (2006) Orthodontic treatment--improvement and standards using the peer assessment rating index. **Angle Orthod**, 76: (2): 260-264.

Onyeaso, C.O. and Begole, E.A. (2007) Relationship between index of complexity, outcome and need, dental aesthetic index, peer assessment rating index, and American Board of Orthodontics objective grading system. **Am J Orthod Dentofacial Orthop**, 131: (2): 248-252.

Pangrazio-Kulbersh, V., Kaczynski, R. and Shunock, M. (1999) Early treatment outcome assessed by the Peer Assessment Rating index. **Am J Orthod Dentofacial Orthop**, 115: (5): 544-550.

Pickering, E.A. and Vig, P. (1975) The occlusal index used to assess orthodontic treatment. **Br J Orthod**, 2: (1): 47-51.

Richmond, S. (2005) **Evaluating Effective Orthodontic Care**. 1st Ed.FIRST Numerics Ltd.

Richmond, S. and Andrews, M. (1993) Orthodontic treatment standards in Norway. **Eur J Orthod**, 15: (1): 7-15.

Richmond, S., Andrews, M. and Roberts, C.T. (1993a) The provision of orthodontic care in the general dental services of England and Wales: extraction patterns, treatment duration, appliance types and standards. **Br J Orthod**, 20: (4): 345-350.

Richmond, S., Shaw, W.C., O'Brien, K.D., *et al.* (1992) The development of the PAR Index (Peer Assessment Rating): reliability and validity. **Eur J Orthod**, 14: (2): 125-139.

Richmond, S., Turbill, E.A. and Andrews, M. (1993b) Calibration of non-dental and dental personnel in the use of the PAR Index. **Br J Orthod**, 20: (3): 231-234.

Roberts, C.T. and Richmond, S. (1997) The design and analysis of reliability studies for the use of epidemiological and audit indices in orthodontics. **Br J Orthod**, 24: (2): 139-147.

Salzmann, J.A. (1968) Handicapping malocclusion assessment to establish treatment priority. **Am J Orthod**, 54: (10): 749-765.

Shaw, W.C., Richmond, S. and O'Brien, K.D. (1995) The use of occlusal indices: a European perspective. **Am J Orthod Dentofacial Orthop**, 107: (1): 1-10.

Stevens, D.R., Flores-Mir, C., Nebbe, B., *et al.* (2006) 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**, 129: (6): 794-803.

Summers, C.J. (1971) The occlusal index: a system for identifying and scoring occlusal disorders. **Am J Orthod**, 59: (6): 552-567.

Summers, C.J. (1972) Tests of validity of indices of occlusion. **Am J Orthod**, 62: (4): 428-429.

Tang and Wei (1990) Assessing treatment effectiveness of removable and fixed orthodontic appliances with the occlusal index. **Am J Orthod Dentofacial Orthop**, 98: (6): 550-556.

Tang and Wei (1993) Recording and measuring malocclusion: A review of the literature. **Am J Orthod Dentofacial Orthop**, 103: (4): 344-351.

Templeton, K.M., Powell, R., Moore, M.B., *et al.* (2006) Are the Peer Assessment Rating Index and the Index of Treatment Complexity, Outcome, and Need suitable measures for orthognathic outcomes? **Eur J Orthod**, 28: (5): 462-466.

Turbill, E.A., Richmond, S. and Andrews, M. (1994) A preliminary comparison of the DPB's grading of completed orthodontic cases with the PAR Index. **Br J Orthod**, 21: (3): 279-285.

Turbill, E.A., Richmond, S. and Wright, J.L. (1996a) Assessment of General Dental Services orthodontic standards: the Dental Practice Board's gradings compared to PAR and IOTN. **Br J Orthod**, 23: (3): 211-220.

Turbill, E.A., Richmond, S. and Wright, J.L. (1996b) A critical assessment of orthodontic standards in England and Wales (1990-1991) in relation to changes in prior approval. **Br J Orthod**, 23: (3): 221-228.

Turbill, E.A., Richmond, S. and Wright, J.L. (1998) A critical assessment of high-earning orthodontists in the General Dental Services of England and Wales (1990-1991). **Br J Orthod**, 25: (1): 47-54.

Turbill, E.A., Richmond, S. and Wright, J.L. (1999) A closer look at General Dental Service orthodontics in England and Wales. II: What determines appliance selection? **Br Dent J**, 187: (5): 271-274.

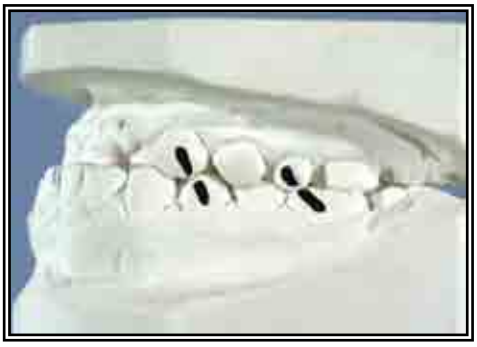
WHO (1966) "An international methodology for epidemiological studies of oral disease". **Epidemiological studies of periodontal disease.Manual No. 5:First draft. Geneva World Health Organization 1966**.

Younis, J.W., Vig, K.W., Rinchuse, D.J., *et al.* (1997) A validation study of three indexes of orthodontic treatment need in the United States. **Community Dent Oral Epidemiol**, 25: (5): 358-362.

APPENDIX

Appendix 1 Case 1 Study Models

Case 1 Pre-treatment

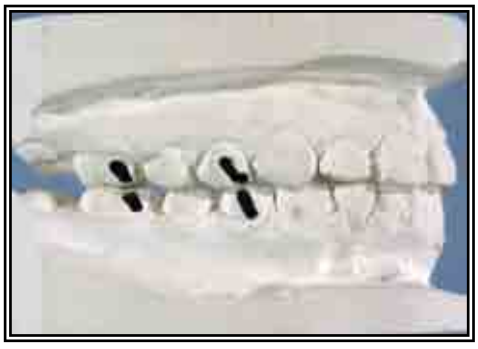


Case 1 Post-treatment

Appendix 2 Case 2 Study Models

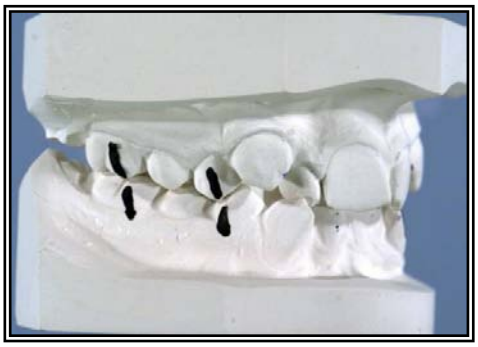
Case 2 Pre-treatment

Case 2 Post-treatment

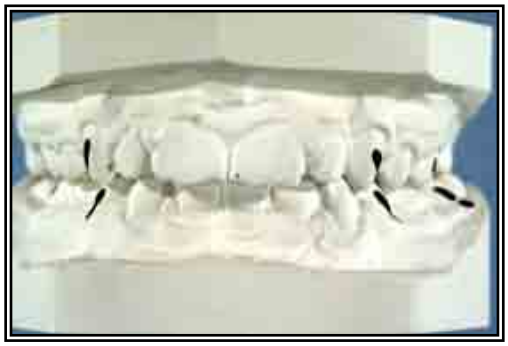


Appendix 3 Case 3 Study Models

Case 3 Pre-treatment



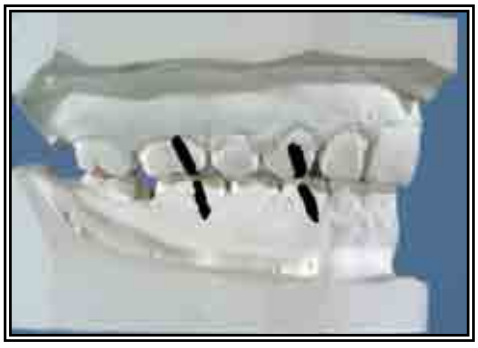
Case 3 Post-treatment



Appendix 4 Case 4 Study Models

Case 4 Pre-treatment

Case 4 Post-treatment



Appendix 5 Data Collection Sheet

UNIVERSITY OF BIRMINGHAM

YEAR OF CALIBRATION / RECALIBRATION: …….

PAR SCORING SHEET

CASE NUMBER		Pre-Treatment		Date									
PAR COMPONENTS	RIGHT									LEFT	UN-WEIGHTED TOTAL	WEIGHTED TOTAL
Upper anterior segments	3-2		2-1		1-1		1-2		2-3			X1
Lower anterior segments	3-2		2-1		1-1		1-2		2-3			X1
Buccal occlusion	Antero-posterior				Right		Left					X1
	Transverse				Right		Left					X1
	Vertical				Right		Left					X1
Overjet	Positive				Negative							X6
Overbite	Overbite				Openbite							X2
Centre line												X4
									TOTAL			

CASE NUMBER		Post-Treatment		Date									
PAR COMPONENTS	RIGHT									LEFT	UN-WEIGHTED TOTAL	WEIGHTED TOTAL
Upper anterior segments	3-2		2-1		1-1		1-2		2-3			X1
Lower anterior segments	3-2		2-1		1-1		1-2		2-3			X1
Buccal occlusion	Antero-posterior				Right		Left					X1
	Transverse				Right		Left					X1
	Vertical				Right		Left					X1
Overjet	Positive				Negative							X6
Overbite	Overbite				Openbite							X2
Centre line												X4
									TOTAL			

ASSESSMENT OF OUTCOME

PAR SCORE		IMPROVEMENT	
Change in PAR score		Greatly improved	
% change in PAR score		Improved	
		Worse or no different	

Appendix 6 Participant Invitation Letter Group 1

UNIVERSITY OF BIRMINGHAM

Re: Calibrated PAR scorers to participate in a study to assess variability in PAR scoring

I am a registrar in Orthodontics and as part of my studies I am undertaking a project to identify and assess differences in PAR scoring between calibrated dental technicians in different regions.

I will compare PAR scores provided by dental technicians in the West Midlands region to those provided by a sample of dental technicins nationally.

I am asking you to be participate as you have been randomly selected from the British Orthodontic Society list of Calibrated PAR Scorers.

If you agree to take part you will be asked to PAR score 8 sets of study models (4 cases pre- and post- treatment). The scores which you provide will be anonymous and your name or contact details will not be recorded. Once all the data has been collected you will receive a copy of the results.

This study will hopefully provide the following useful information:

- How PAR scorers in the West Midlands compare to PAR scorers nationally
- Consistency in PAR scoring between calibrated PAR scorers
- Specific Componets of the PAR index causing variability between PAR scorers

I hope you agree to participate in this study and should you have any questions please do not resitate to contact me. (Email [redacted])

This project is being supervised by Dr. A. Dhopatkar, Head of Orthodontics, School of Dentistry, Birmingham.

Many thanks

Yours sincerely

Miss Criona Harte
Specialist Registrar in Orthodontics

Appendix 7 Participant Invitation Letter Group 2

UNIVERSITY OF BIRMINGHAM

Re: Calibrated PAR scorers to participate in a study to assess variability in PAR scoring

I am a registrar in Orthodontics and as part of my studies I am undertaking a project to identify and assess differences in PAR scoring between calibrated dental technicians in different regions.

I will compare PAR scores provided by dental technicians in the West Midlands region to those provided by a sample of dental technicians nationally.

If you agree to take part you will be asked to PAR score 8 sets of study models (4 cases pre- and post- treatment). The scores which you provide will be anonymous and your name or contact details will not be recorded. Once all the data has been collected you will receive a copy of the results.

This study will hopefully provide the following useful information:

- How PAR scorers in the West Midlands compare to PAR scorers nationally
- Consistency in PAR scoring between calibrated PAR scorers
- Specific Componets of the PAR index causing variability between PAR scorers

I hope you agree to participate in this study and should you have any questions please do not hesitate to contact me. (Email [redacted])

This project is being supervised by Dr. A. Dhopatkar, Head of Orthodontics, School of Dentistry, Birmingham.

Many thanks

Yours sincerely

Miss Criona Harte
Specialist Registrar in Orthodontics

Appendix 8 Raw Data Total Weighted PAR scores

Examiner	Pre 1	Post 1	Pre 2	Post 2	Pre 3	Post 3	Pre 4	Post 4	Group
1	41	6	52	9	69	21	38	4	1
2	37	3	45	3	47	11	36	2	1
3	41	7	31	3	68	10	37	4	1
4	35	6	32	3	55	11	37	5	1
5	39	5	34	2	64	12	38	4	1
6	37	7	41	5	58	11	32	6	1
7	35	5	43	4	58	8	42	4	1
8	35	4	36	3	54	8	38	5	1
9	36	5	36	6	60	7	41	4	1
10	38	5	42	4	57	14	43	5	1
11	36	4	37	7	57	14	45	3	1
12	39	4	43	2	60	8	20	3	1
13	34	5	31	3	51	8	38	5	1
14	38	5	38	7	35	11	43	5	1
15	32	5	35	4	58	11	44	4	2
16	36	4	42	3	41	8	41	4	2
17	34	4	36	3	64	19	39	4	2
18	39	8	40	6	59	15	35	7	2
19	39	8	39	6	59	15	35	7	2
20	38	8	39	5	59	13	35	6	2
21	36	7	40	6	59	15	35	7	2
22	38	7	35	4	58	12	38	4	2
23	55	5	41	2	45	18	35	3	2
24	38	7	39	10	59	15	35	6	2
25	38	6	44	8	68	11	40	4	2
26	36	4	45	2	60	10	42	4	2
27	36	6	45	4	59	10	42	5	2
28	33	4	54	1	43	1	33	4	2
*	16*	3	37	4	52	9	34	5	1

*One anomalous score was obtained. Following statistical advice data from this examiner was not included in the statistical analysis plan. On the whole data normality assumption was not violated.

Appendix 9 Component scores Case 1 Pre-treatment

Examiner	ULS	LLS	BUCCAL	OVERJET	OVERBITE	CENTRELINE
1	5	1	7	24	4	0
2	4	1	6	24	2	0
3	4	2	7	24	4	0
4	5	0	6	24	0	0
5	4	1	6	24	4	0
6	5	1	7	24	0	0
7	4	1	6	24	0	0
8	5	1	5	24	0	0
9	5	2	5	24	0	0
10	4	1	7	24	2	0
11	5	1	6	24	0	0
12	5	1	7	24	2	0
13	4	0	6	24	0	0
14	5	2	7	24	0	0
15	4	1	3	24	0	0
16	5	1	6	24	0	0
17	3	1	6	24	0	0
18	4	0	9	24	2	0
19	4	0	9	24	2	0
20	4	0	8	24	2	0
21	4	1	7	24	0	0
22	5	2	5	24	2	0
23	6	2	7	36	4	0
24	4	1	7	24	2	0
25	4	1	7	24	2	0
26	4	0	6	24	2	0
27	4	0	6	24	2	0
28	3	1	5	24	0	0

Appendix 10 Component scores Case 1 Post-treatment

Examiner	ULS	LLS	BUCCAL	OVERJET	OVERBITE	CENTRELINE
1	0	0	6	0	0	0
2	0	0	3	0	0	0
3	0	0	7	0	0	0
4	0	0	6	0	0	0
5	0	0	5	0	0	0
6	0	0	7	0	0	0
7	0	0	5	0	0	0
8	0	0	4	0	0	0
9	0	0	5	0	0	0
10	0	0	5	0	0	0
11	0	0	4	0	0	0
12	0	0	4	0	0	0
13	0	0	5	0	0	0
14	1	0	4	0	0	0
15	2	0	3	0	0	0
16	0	0	4	0	0	0
17	0	0	4	0	0	0
18	0	0	8	0	0	0
19	0	0	8	0	0	0
20	0	0	8	0	0	1
21	0	0	7	0	0	0
22	1	0	6	0	0	0
23	0	0	5	0	0	0
24	0	0	7	0	0	0
25	1	1	4	0	0	0
26	0	0	4	0	0	0
27	2	0	4	0	0	0
28	0	0	4	0	0	0

Appendix 11 Component scores Case 2 Pre-treatment

Examiner	ULS	LLS	BUCCAL	OVERJET	OVERBITE	CENTRELINE
1	7	4	3	30	0	8
2	4	4	3	24	2	8
3	5	3	3	18	2	0
4	6	3	3	18	2	0
5	4	2	2	18	0	8
6	6	3	4	24	0	4
7	8	3	2	24	2	4
8	8	4	2	18	0	4
9	8	5	1	18	0	4
10	8	3	3	24	0	4
11	7	5	3	18	0	4
12	8	5	2	18	2	8
13	5	2	2	18	0	4
14	7	4	5	18	0	4
15	6	3	2	18	2	4
16	9	3	2	24	0	4
17	7	3	2	18	2	4
18	5	3	4	24	0	4
19	5	3	4	24	0	4
20	5	3	3	24	0	4
21	5	3	4	24	0	4
22	6	3	2	18	2	4
23	7	4	2	24	4	0
24	5	2	4	24	0	4
25	8	4	2	24	6	0
26	4	0	6	24	2	0
27	5	4	2	24	2	8
28	5	4	1	24	4	16

Appendix 12 Component scores Case 2 Post-treatment

Examiner	ULS	LLS	BUCCAL	OVERJET	OVERBITE	CENTRELINE
1	0	0	3	6	0	0
2	1	0	2	0	0	0
3	0	0	3	0	0	0
4	0	0	3	0	0	0
5	0	0	2	0	0	0
6	1	0	4	0	0	0
7	2	0	2	0	0	0
8	0	0	3	0	0	0
9	2	0	4	0	0	0
10	1	0	3	0	0	0
11	1	0	2	0	0	4
12	0	0	2	0	0	0
13	1	0	2	0	0	0
14	2	1	4	0	0	0
15	2	0	2	0	0	0
16	1	0	2	0	0	0
17	1	0	2	0	0	0
18	1	0	5	0	0	0
19	1	0	5	0	0	0
20	0	4	0	0	0	0
21	1	0	5	0	0	0
22	1	0	3	0	0	0
23	0	0	2	0	0	0
24	0	0	4	6	0	0
25	1	1	4	0	2	0
26	1	1	0	0	0	0
27	1	1	2	0	0	0
28	0	0	1	0	0	0

Appendix 13 Component scores Case 3 Pre-treatment

Examiner	ULS	LLS	BUCCAL	OVERJET	OVERBITE	CENTRELINE
1	8	12	7	30	4	8
2	8	13	2	12	4	8
3	11	18	3	30	2	4
4	6	10	3	30	2	4
5	5	12	5	30	4	8
6	6	11	5	30	2	4
7	6	10	2	30	2	8
8	6	10	2	30	2	4
9	7	12	5	30	2	4
10	6	11	2	30	4	4
11	8	10	3	30	2	4
12	6	10	4	30	2	8
13	7	9	5	24	2	4
14	6	11	6	6	2	4
15	6	10	0	36	2	4
16	6	10	1	12	4	8
17	8	10	2	36	4	4
18	6	10	7	30	2	4
19	6	10	7	30	2	4
20	6	10	7	30	2	4
21	6	10	7	30	2	4
22	6	10	2	30	6	4
23	8	11	2	12	4	8
24	6	10	1	36	2	4
25	6	13	3	36	2	8
26	6	11	3	30	6	4
27	6	10	3	30	6	4
28	10	9	6	6	4	8

Appendix 14 Component scores Case 3 Post-treatment

Examiner	ULS	LLS	BUCCAL	OVERJET	OVERBITE	CENTRELINE
1	0	0	5	6	2	8
2	0	0	3	6	2	0
3	0	0	4	6	0	0
4	0	0	3	6	2	0
5	0	0	4	6	2	0
6	0	0	5	6	0	0
7	0	0	2	6	0	0
8	0	0	2	6	0	0
9	0	0	1	6	0	0
10	0	0	4	6	0	4
11	0	0	2	6	2	4
12	0	0	2	6	0	0
13	0	0	2	6	0	0
14	0	0	3	6	2	0
15	0	1	2	6	2	0
16	0	0	2	6	0	0
17	2	1	2	12	2	0
18	0	0	7	6	2	0
19	0	0	7	6	2	0
20	0	7	6	2	0	0
21	0	0	7	6	2	0
22	1	0	3	6	2	0
23	0	0	2	12	4	0
24	0	0	7	6	2	0
25	1	1	1	6	2	0
26	0	1	1	6	2	0
27	0	1	1	6	2	0
28	0	1	0	0	0	0

Appendix 15 Component scores Case 4 Pre-treatment

Examiner	ULS	LLS	BUCCAL	OVERJET	OVERBITE	CENTRELINE
1	6	4	2	24	2	0
2	8	4	2	18	4	0
3	11	4	2	18	2	0
4	6	4	3	18	6	0
5	5	4	1	24	4	0
6	6	4	4	18	0	0
7	6	4	2	24	6	0
8	6	4	2	24	2	0
9	10	5	2	18	6	0
10	7	4	2	24	6	0
11	6	5	4	24	2	4
12	7	5	2	6	0	0
13	8	4	2	18	2	4
14	7	5	3	24	4	0
15	8	4	2	24	6	0
16	7	4	2	24	4	0
17	8	3	2	18	8	0
18	6	4	5	18	2	0
19	6	4	5	18	2	0
20	6	4	5	18	2	0
21	6	4	5	18	2	0
22	7	4	3	18	2	4
23	8	4	1	18	4	0
24	6	4	1	18	6	0
25	5	3	2	24	2	4
26	8	4	2	24	4	0
27	8	4	2	24	4	0
28	6	4	1	18	4	0

Appendix 16 Component scores Case 4 Post-treatment

Examiner	ULS	LLS	BUCCAL	OVERJET	OVERBITE	CENTRELINE
1	1	0	3	0	0	0
2	0	0	2	0	0	0
3	0	0	4	0	0	0
4	0	0	5	0	0	0
5	0	0	4	0	0	0
6	1	0	5	0	0	0
7	1	0	3	0	0	0
8	1	0	4	0	0	0
9	0	0	4	0	0	0
10	1	0	4	0	0	0
11	0	0	3	0	0	0
12	0	0	3	0	0	0
13	1	0	4	0	0	0
14	1	0	4	0	0	0
15	1	0	3	0	0	0
16	0	0	4	0	0	0
17	1	0	3	0	0	0
18	0	0	7	0	0	0
19	0	0	7	0	0	0
20	0	0	6	0	0	0
21	0	0	7	0	0	0
22	0	0	4	0	0	0
23	0	0	3	0	0	0
24	0	0	6	0	0	0
25	1	1	2	0	0	0
26	1	0	3	0	0	0
27	1	0	4	0	0	0
28	1	0	3	0	0	0

Appendix 17 Year of Calibration/Recalibration

Examiner	Calibration	Recalibration
1	2001	
2	2001	
3	2000	2008
4	2006	
5	1993	
6	2001	2008
7	2003	
8	2008	
9	2009	
10	1998	
11	2008	
12	2008	
13	2004	
14	2001	
15	2003	
16	1992	
17	2001	
18	2001	
19	1997	
20	1998	
21	2005	
22	2006	
23	2000	
24	2002	
25	2000	
26	2009	
27	2000	
28	2001	
*	1996	

Appendix 18 PAR Outcome Categories

Examiner	Case 1	Case 2	Case 3	Case 4
1	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
2	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
3	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
4	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
5	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
6	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
7	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
8	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
9	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
10	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
11	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
12	Greatly Improved	Greatly Improved	Greatly Improved	Improved
13	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
14	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
15	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
16	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
17	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
18	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
19	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
20	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
21	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
22	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
23	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
24	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
25	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
26	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
27	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved
28	Greatly Improved	Greatly Improved	Greatly Improved	Greatly Improved