A comparison of orthodontic treatment outcomes using the Objective Grading System (OGS) and the Peer Assessment Rating (PAR) index

Orfan Chalabi,[*] Charles Brian Preston,[†] Thikriat S. Al-Jewair[†] and Sawsan Tabbaa[†]
Private practice, Orange County, California[*] and Department of Orthodontics, State University of New York, New York,[†] USA

Introduction: The use of objective criteria is essential to uniformly quantify and measure the severity of malocclusions and the efficacy of different treatment modalities. The Peer Assessment Rating (PAR) index and, more recently, the American Board of Orthodontics Objective Grading System (OGS) were developed to fulfill this need.
Aim: The aim of this retrospective study was to assess and compare treatment outcomes using the UK and US weighted PAR and the OGS.
Materials and methods: The sample consisted of randomly selected records of 50 patients treated by residents in one postgraduate orthodontic clinic. UK and US weightings for the PAR index were applied and compared with OGS.
Results: There was no statistically significant association between the OGS and the PAR index grading systems. Neither the UK nor the US PAR weightings showed statistically significant correlation with the OGS. All cases were 'greatly improved' or 'improved' according to the PAR index, while most cases (62%) failed according to OGS. There was a statistically significant correlation between the unweighted PAR index and the OGS ($r = -0.32$, $p = 0.024$). The US and the UK weightings for the PAR were highly correlated ($r = 0.90$, $p < 0.001$). Both weighting systems were also highly correlated with the unweighted PAR ($p < 0.001$). There were no gender differences found in any of the scoring systems.
Conclusions: The current PAR index cannot replace the OGS for evaluating treatment outcomes. The current OGS cannot detect the improvement achieved in a treated case.
(Aust Orthod J 2015; 31: 157-164)

Orfan Chalabi: orfanchalabi@yahoo.com; Charles Brian Preston: cbp@buffalo.edu; Thikriat S. Al-Jewair: thikriat@buffalo.com;
Sawsan Tabbaa: tabbaa@buffalo.edu

Introduction

Orthodontic treatment results are most often subjectively graded but may be assessed by objective methods, either in clinical settings, study groups, national or state board examinations.[1] Several indices have been used to impartially evaluate a malocclusion and treatment outcome.[1,2]

The American Board of Orthodontics (ABO) has published the Objective Grading System (OGS) to evaluate treatment outcomes using post-treatment dental casts and panoramic radiographs.[3] The OGS scores eight occlusal traits and produces a total score. An increasingly higher score indicates a worsening outcome. In general, a case that scores more than 30 points will fail assessment, while a case that scores less than 20 points will pass.

Another popular assessment tool is the Peer Assessment Rating (PAR) index,[4] which measures occlusal traits and allocates scores for tooth alignment, dental impaction, relationships of the buccal segments, overjet, overbite and midline discrepancies. These determinations are made on the pre- and post-

treatment dental casts, and differences in the resultant PAR scores reflect the effect of treatment.[5] Weightings are assigned to each component of the PAR index to reflect their relative importance and to produce a weighted PAR index score.[6] The greater the mean percentage reduction in the weighted PAR score, the greater the degree of orthodontic improvement achieved.[4] The mean percent reduction in the PAR scores ranged from 68% to 78% in several studies.[7-9]

The PAR index has been validated in the United Kingdom (PAR UK).[5,10] The index has also been applied in the United States (PAR US), but using different weightings and by eliminating the mandibular anterior alignment component.[11] McKnight et al.[12] examined records of 27 patients who had been recalled an average of nine years after the completion of a two-stage (functional/fixed) Class II treatment. Using the UK and US weightings, minimal differences were found between the two systems; however, the nature of observed relapse argued against the American exclusion of the lower labial segment.

The aim of this retrospective study was therefore to assess and compare the treatment outcomes using the PAR index (UK and US weightings), the OGS and to assess gender differences.

Materials and methods

A sample size of 50 was considered statistically adequate to detect a significant difference between pre- and post-treatment PAR measurements and between the PAR and OGS systems.

Fifty subjects representing a mixture of malocclusion types, ethnic groups and ages were randomly selected from the Postgraduate Orthodontic Clinic at the University at Buffalo. The inclusion criteria were permanent dentitions at pretreatment and subjects treated with multi-bracket fixed upper and lower edgewise appliances (0.018 inch × 0.025 inch slot), the availability of pre- and post-treatment study casts without attached appliances, and a final panoramic radiograph. The exclusion criteria were cases treated by orthognathic surgery, broken casts, and those that only represented phase I treatment. One calibrated and PAR assessment certified investigator (O.C.) measured each set of study casts and panoramic radiographs.

The OGS measurements

An ABO kit with a metal gauge was used to examine every case. The width of the gauge was 0.5 mm and its height was 1 mm. The OGS comprised eight components: tooth alignment, marginal ridge heights, bucco-lingual tooth inclination, occlusal contacts, occlusal relationships, overjet, interproximal contacts and root angulation.[3] After a total number of case points was calculated, a case that lost more than 30 points was considered a failure. A case that lost less than 20 points was considered to have passed. A case that lost 20–30 points was considered a 'maybe' (borderline).

The PAR index measurements

All PAR index measurements were made on the pre- and post-treatment records using the PAR index ruler. The measurements included: tooth alignment, buccal occlusion, overjet, overbite and centre line.[4]

After generating a total score of the various components of the index, the UK and the US weightings were applied (Table I). The improvement in the PAR scores (success of treatment) was categorised into: 'greatly

Table I. UK and US PAR weightings.

PAR component	UK study	US study
Upper labial segment alignment	1	1
Lower labial segment alignment	1	0
Overbite	2	3
Overjet	6	5
Midline	4	3
Right buccal segment relationship	1	2
Left buccal segment relationship	1	2

improved' if the percentage reduction in PAR index was at least 30% and there was >22 points reduction in the score; 'improved' if there was a percentage reduction of at least 30% and <22 points reduction on the score; and 'worse or no difference'.

The PAR index measurements were completed first, and after an interval of one month the OGS measurements were determined.

Intra-examiner reliability

Sixteen sets of records were randomly selected and remeasured by one investigator after an interval of two weeks. The measurements were performed according to the PAR UK and the OGS. The error of the method was estimated by using Dahlberg's formula.[13] The agreement between measurements was assessed using the intra-class correlation coefficient (ICC).

Statistical analysis

Data were analysed using SPSS version 11 for Windows (SPSS Inc., IL, USA). Descriptive statistics were initially conducted for all variables. The relationships between the PAR and OGS scoring methods and differences between males and females were assessed

using Chi-square tests for categorical variables and *t*-tests or one-way ANOVA, as appropriate. The correlations between the scoring methods were determined using Pearson's correlation coefficient. All statistical testing was two-tailed at the 5% level of significance.

Results

Fifty sets of models from 31 females (62%) and 19 males (38%) were included. The results of the error of the method showed that the standard deviations were small relative to the measurement scales (Table II). The pre- and post-treatment UK weighted PAR scores ranged from 11 to 51 points and 0 to 19 points respectively. The percentage improvement in treatment outcome ranged from 50% to 100%. The assessment of the agreement between the first and second round of measurements indicated a high level of agreement.

OGS results

The majority of cases (62%) were classified as failures and only a small percentage (8%) classified as a pass (Table III).

Table II. Intra-examiner reliability analysis.

	Mean absolute difference	Dahlberg standard deviation	ICC
OGS	2.44	2.08	0.947
PAR UK pretreatment	0.88	0.79	0.994
PAR UK post-treatment	0.50	0.56	0.976
% PAR Improvement	2.58	2.90	0.950

ICC: intra-class correlation coefficient

Table III. Sample description for OGS and PAR scoring systems (N = 50).

OGS			UK weight				US weight		
Score	Category	N (%)	PAR category	N (%)	Mean %	SD	N (%)	Mean %	SD
≥31	Generally fail	31 (62)	Improved	21 (42)	79.50	11.95	23 (46)	75.42	15.81
21–30	Maybe	15 (30)	Greatly improved	29 (58)	88.24	9.05	27 (54)	85.85	10.75
≤20	Generally pass	4 (8)	Total	50 (100)	84.62	11.16	50 (100)	81.05	14.18

PAR results

The overall mean percentage improvement was 84.62% and 81.05% according to the UK and US weightings, respectively (Table III). All patients were judged to have 'improved' and the majority of cases 'greatly improved'. Four of the 'greatly improved' in the PAR UK weighting were judged to be 'improved' by the PAR US weighting and two of the 'improved' in the PAR UK weighting were judged to be 'greatly improved' by the PAR US weighting, which was a net shift of two cases.

Comparison of the OGS and the PAR

There were no statistically significant associations between the OGS and the PAR index (p = 0.738) (Table IV). Overall, 62% (31/50) of the sample received an OGS score of ≥31. Of the PAR UK 'improved' group (13/21), 62% failed according to the OGS evaluation, and the same percentage (62%) of the PAR UK 'greatly improved' group (18/29) also failed the OGS. Fifty-eight percent of the sample (29/50) were judged as 'greatly improved' by the PAR UK index, yet 18 out of 31 patients (58%) in this group received an OGS score of ≥31. A similar comparison using the US, instead of the UK weighting, revealed no significant association between the two grading systems (p = 0.517).

The results of the percentage improvement score for either weighting systems of the PAR in association with the OGS are presented in Table V. The results of the one-way analysis of variance displayed no statistically significant relationships for the PAR UK in comparison with the OGS (F = 0.01; df = 2,47; p = 0.988), or for the PAR US compared with the OGS (F = 0.17; df = 2,47; p = 0.845).

Alternatively, it was possible to evaluate the mean score from the OGS with respect to the grades assigned by the PAR weighted systems (Table VI). There were no statistically significant relationships between the patients' mean OGS scores with respect to their grades on the PAR UK (t-statistic = 0.030, p = 0.976) and US (t-statistic = -0.68, p = 0.497) systems.

Table VII presents the correlations between the OGS scores and the percentage improvement in the PAR scores. The correlations involving the OGS were negative, which indicated that for smaller OGS scores (passing grade) the percentage improvement in the PAR score was larger (improving). However, the correlations between OGS and PAR were weak and not statistically significant for either weighting system, but statistically significant (p = 0.024) for the unweighted PAR scores. The percentage improvements in the UK and the US PAR scores were significantly correlated with each other (p < 0.001).

Table IV. PAR UK* and US** scorings versus OGS scoring.

PAR categories	Scoring System	OGS categories (N)			
		≥31	21–30	≤20	Total
Improved	UK	13	7	1	21
	US	16	6	1	23
Greatly improved	UK	18	8	3	29
	US	15	9	3	27
	Total	31	15	4	50

* Chi-square = 0.61, df = 2, p-value = 0.738
** Chi-square = 1.32, df = 2, p-value = 0.517

Table V. Mean percentage improvement of the PAR for the OGS Scores.

OGS	N	UK weight		US weight	
		Mean %	SD	Mean %	SD
≥31	31	84.78	10.45	81.96	12.14
21–30	15	84.24	12.83	79.35	17.48
≤20	4	84.77	13.15	80.45	18.98
Overall	50	84.62	11.16	81.05	14.18

An assessment of the statistical association between gender and the three scoring systems indicated no statistically significant differences for any of the three scoring systems (Table VIII). Of the 31 patients who received an OGS score of over 30, 64.5% (20) were females, a percentage almost identical to the overall percentage of the females included in the sample (62%). A similar result was found when gender was compared between the percentage improvements in PAR scores based on the UK weighting system (*p* = 0.50).

Discussion

The aim of the present study was to assess and compare treatment outcomes using the UK and US weighted PAR and the OGS indices. The results showed the mean percentage improvement was 84.62% ± 11.16% and 81.05% ± 14.18% for the PAR UK and US weightings, respectively. The findings agree with Richmond,[14] and suggest that a good standard of treatment was achieved. Richmond believed that a percentage reduction in weighted PAR score greater than 70% represented a significant improvement in the standard of an occlusion. Twenty-nine cases (58%) of the sample of patients included in the present study were 'greatly improved' as a result of orthodontic treatment and twenty-one (42%) were 'improved'.

The current results are similar to the findings of Dyken et al.,[15] who compared American Board of Orthodontics (ABO) cases with those treated by graduate students at the University of Alabama. The mean percentage reduction for the ABO cases was 87.9%, and for the graduate students' cases it was 81.7%. The findings are higher than two previous studies,[9,16] which reported a mean percentage improvement of 68.9% and 78%, respectively.

Table VI. Mean OGS scores for the PAR index.

PAR categories	UK weight	OGS		US weight	OGS	
	N	Mean	SE	N	Mean	SE
Improved	21	32.10	1.54	23	33.09	1.42
Greatly improved	29	32.17	1.91	27	31.33	2.03
t-statistic		0.02970			-0.68476	
p-value*		0.9764			0.4968	

* Significant at 0.05

Table VII. Correlations among OGS and percentage improvement in PAR Scores for UK weights, US weights and unweighted scores.

Grading systems	Correlation	*p*-values*
OGS vs. %PAR UK weight	-0.16	0.254
OGS vs. %PAR US weight	-0.11	0.450
OGS vs. %PAR unweighted	-0.32	0.024
%PAR UK weight vs. %PAR US weight	0.90	<0.001
%PAR UK weight vs. %PAR unweighted	0.87	<0.001
%PAR US weight vs. %PAR unweighted	0.73	<0.001

* Significant at 0.05

Table VIII. Differences between males and females for OGS scores and percentage improvement in PAR scores for both UK and US weight systems.

Scoring method	Male		Female			
	Mean	SE	Mean	SE	t-statistic	*p*-value*
OGS	31.32	2.20	32.65	1.56	0.50	0.616
PAR UK weight	82.33	3.17	86.02	1.64	1.14	0.261
PAR US weight	79.52	3.98	82.00	2.17	0.60	0.554

* Significant at 0.05

All cases in the present study were 'improved' or 'greatly improved', yet 62% failed the OGS and only 8% would pass with certainty.

There were no statistically significant associations between the PAR UK and the OGS indices (p = 0.73), and the correlation was negative (r = -0.16). This implies that predictions cannot be made for the possible percentage improvement in the PAR index from the OGS scores. There were no statistically significant differences observed between the OGS mean scores and the scores achieved using the PAR index. The same tests were applied using the American weighting for the PAR index (PAR US) and similar findings were found using the PAR US and the PAR UK. Both weighing systems were not significantly associated with the OGS. However, the results of the PAR UK revealed a better correlation with the OGS scores. It should be cautioned that since there were no statistically significant associations between either of the PAR systems and the OGS results, it is not possible to favour the PAR UK over the PAR US index. Other studies have noted the deficiencies in the PAR US weighting of evaluation. McKnight et al.[12] argued in favour of the PAR UK over the PAR US because the latter index excluded the lower labial segment alignment. Correspondingly, Dyken et al.[15] were unable to use the PAR US index in their statistics because of a need to include the lower anterior alignment in their studies, as they considered this odontometric character highly important. Dyken et al. compared the PAR US and the PAR UK indices and determined that the recorded percentage improvements were significantly associated with each other (p < 0.001). Previous research[12] found that the differences between the scores generated by the PAR UK and the PAR US weightings were relatively small.

The present study showed a minimal statistical significance between the unweighted PAR index and the OGS (p = 0.024). The result may indicate that by modifying the weighting of the PAR index, a correlation may be reached between the OGS and the PAR. Alternatively, changes can be made to the OGS to include pretreatment scores and this may help to measure the obtained improvement, as well as providing precision in the detection of the deficiency of the achieved occlusion.

It may be argued that the PAR index is insufficiently precise. Buchanan et al.[17] stated that the PAR index had shortcomings in that it failed to adequately record features such as incisor torque, posterior alignment and changes in arch dimensions. Richmond et al.[4] indicated that the minor deviations from normal cannot be 'greatly improved', if a case was not severe enough in the first instance. Fox[18] noted that any case that had an overjet of 9 mm or greater scored the same for this occlusal parameter based on the PAR index. However, many studies suggested that the sensitivity of the PAR index was sufficient to detect differences in treatment outcome[7,10,15,17-19] and may also be used to discriminate between the results obtained by different orthodontists. Shaw et al.[20] used the PAR index for this purpose when cases treated in England were compared with those treated in Wales. Similarly, the PAR index has been used to evaluate treatment results obtained by orthodontists in different countries.[9]

A great improvement was achieved in the present sample when assessed by the PAR index. The OGS only examines a final outcome and cannot detect an improvement brought about by a particular treatment modality. This shortcoming of the OGS was also noted by Yang-Powers et al.,[21] who stated that the ABO OGS only defined treatment outcome and did not take into account the severity of the original malocclusion or the difficulty of treatment. When evaluating the efficacy or effectiveness of orthodontic treatment, a change obtained using a particular scoring system from the pretreatment to the post-treatment stage was an important consideration.[22]

In general, the clinical use of the terms 'failure' or 'treatment objectives not achieved' are considered relative. Because a considerable improvement takes place during treatment in most cases, the term 'partial success' might be more appropriate.

Past studies have identified limitations associated with PAR index scoring and have indicated that the problems relate mainly to the generic weightings given to the occlusal traits of overjet and overbite. The relatively high weighting assigned to overjet may influence the index to an extent that it is unduly sensitive to any malocclusion in which an overjet is increased.[19] Further limitations of the PAR index are that occlusions with initial scores of less than 22 points cannot become 'greatly improved' by treatment,[6] and that changes in cephalometric parameters that reflect the skeletal components of a malocclusion are not considered in the quantitative evaluation of the PAR Index.[5]

The PAR and the OGS can be considered as mechanical systems of measurement that are incapable of evaluating all orthodontic treatment outcomes. Previous authors describe occlusal indices as measures of orthodontic outcomes.[1,5,14,18,23,24] Changes in facial profile or cephalometric parameters that reflect the skeletal component of malocclusion are not considered in the quantitative evaluation. Unfortunately, the measurement of these important variables by valid and reliable methods has not been achieved. This has been attributed to individual biologic variation, which requires discrimination between changes produced by orthodontic intervention compared with those caused by growth and development of the facial complex. In addition, the ideal cephalometric analysis or cephalometric goals of orthodontic treatment are controversial and no consensus exists within the orthodontic profession. Thirdly, no universally accepted methods currently exist to assess changes in facial profile as an outcome measure.

In the present study, no differences were found between male and female outcomes following OGS or PAR score assessment. These findings correspond with past investigators who reported no correlation between gender and the changes recorded in PAR scores.[25,26]

The present study identified weaknesses and strengths related to the two applied indices. A modification of the indices is recommended in order to provide a more objective insight into the provided orthodontic treatment. This might be managed by an adjustment of the PAR weighting to reduce the emphasis on overjet and stricter criteria for tooth alignment. In addition, modifying the OGS to include the achieved improvement in a treated case would be beneficial, as it is an important component of orthodontic treatment.

Conclusions

- The mean PAR UK score reduction was 84.62%, and for the PAR US it was 81.05%. This indicated a good standard of orthodontic treatment.

- Sixty-two percent of the cases would have failed the ABO examination according to the OGS.

- There was no statistically significant association between the OGS and the weighted PAR index. The OGS, however, was significantly associated with the unweighted PAR.

- There were no statistically significant gender differences evident between the scoring systems.

- The current PAR cannot replace the OGS for evaluating the American Board of Orthodontics cases. The current OGS cannot detect an improvement achieved in a treated case.

Corresponding author

Orfan Chalabi
1501 E. Katella Ave.
Orange County
California 92867
USA

Email: orfanchalabi@yahoo.com

References

1. Pangrazio-Kulbersh V, Kaczynski R, Shunock M. Early treatment outcome assessed by the Peer Assessment Rating index. Am J Orthod Dentofacial Orthop 1999;115:544-50.
2. Summers CJ. The occlusal index: a system for identifying and scoring occlusal disorders. Am J Orthod 1971;59:552-67.
3. Casko JS, Vaden JL, Kokich VG, Damone J, James RD, Cangialosi TJ et al. Objective grading system for dental casts and panoramic radiographs. American Board of Orthodontics. Am J Orthod Dentofacial Orthop 1998;114:589-99.
4. Richmond S, Shaw WC, Roberts CT, Andrews M. The PAR Index (Peer Assessment Rating): methods to determine outcome of orthodontic treatment in terms of improvement and standards. Eur J Orthod 1992;14:180-7.
5. DeGuzman L, Bahiraei D, Vig KW, Vig PS, Weyant RJ, O'Brien K. The validation of the Peer Assessment Rating index for malocclusion severity and treatment difficulty. Am J Orthod Dentofacial Orthop 1995;107:172-6.
6. Hamdan AM, Rock WP. An appraisal of the Peer Assessment Rating (PAR) Index and a suggested new weighting system. Eur J Orthod 1999;21:181-92.
7. O'Brien KD, Shaw WC, Roberts CT. The use of occlusal indices in assessing the provision of orthodontic treatment by the hospital orthodontic service of England and Wales. Br J Orthod 1993;20:25-35.
8. O'Brien KD, Robbins R, Vig KW, Vig PS, Shnorhokian H, Weyant R. The effectiveness of Class II, division 1 treatment. Am J Orthod Dentofacial Orthop 1995;107:329-34.
9. Richmond S, Andrews M. Orthodontic treatment standards in Norway. Eur J Orthod 1993;15:7-15.
10. Richmond S, Shaw WC, O'Brien KD, Buchanan IB, Jones R, Stephens CD et al. The development of the PAR Index (Peer Assessment Rating): reliability and validity. Eur J Orthod 1992;14:125-39.
11. Firestone AR, Beck FM, Beglin FM, Vig KW. Evaluation of the peer assessment rating (PAR) index as an index of orthodontic treatment need. Am J Orthod Dentofacial Orthop 2002;122:463-9.
12. McKnight MM, Daniels CP, Johnston LE Jr. A retrospective study of two-stage treatment outcomes assessed with two modified PAR indices. Angle Orthod 1998;68:521-4; discussion 525-6.
13. Dahlberg G. Statistical methods for medical and biological students. New York: Interscience Publications, 1940.

14. Richmond S. Personal audit in orthodontics. Br J Orthod 1993;20:135-44.

15. Dyken RA, Sadowsky PL, Hurst D. Orthodontic outcomes assessment using the peer assessment rating index. Angle Orthod 2001;71:164-9.

16. al Yami EA, Kuijpers-Jagtman AM, van 't Hof MA. Occlusal outcome of orthodontic treatment. Angle Orthod 1998;68:439-44.

17. Buchanan IB, Russell JI, Clark JD. Practical application of the PAR index: an illustrative comparison of the outcome of treatment using two fixed appliance techniques. Br J Orthod 1996;23:351-7.

18. Fox NA. The first 100 cases: a personal audit of orthodontic treatment assessed by the PAR (peer assessment rating) index. Br Br Dent J 1993;174:290-7.

19. Kerr WJ, Buchanan IB, McColl JH. Use of the PAR index in assessing the effectiveness of removable orthodontic appliances. Br J Orthod 1993;20:351-7.

20. Shaw WC, Richmond S, O'Brien KD, Brook P, Stephens CD. Quality control in orthodontics: indices of treatment need and treatment standards. Br Dent J 1991;170:107-12.

21. Yang-Powers LC, Sadowsky C, Rosenstein S, BeGole EA. Treatment outcome in a graduate orthodontic clinic using the American Board of Orthodontics grading system. Am J Orthod Dentofacial Orthop 2002;122:451-5.

22. Pickering EA, Vig P. The occlusal index used to assess orthodontic treatment. Br J Orthod 1975;2:47-51.

23. Tahir E, Sadowsky C, Schneider BJ. An assessment of treatment outcome in American Board of Orthodontics cases. Am J Orthod Dentofacial Orthop 1997;111:335-42.

24. Ngan P, Yiu C. Evaluation of treatment and posttreatment changes of protraction facemask treatment using the PAR index. Am J Orthod Dentofacial Orthop 2000;118:414-20.

25. Birkeland K, Furevik J, Bøe OE, Wisth PJ. Evaluation of treatment and post-treatment changes by the PAR Index. Eur J Orthod 1997;19:279-88.

26. Taylor PJ, Kerr WJ, McColl JH. Factors associated with the standard and duration of orthodontic treatment. Br J Orthod 1996;23:335-41.