SPECIAL ARTICLE

Posttreatment tooth movement: For better or for worse

Peter M. Greco,[a] Jeryl D. English,[b] Barry S. Briss,[c] Scott A. Jamieson,[d] Marvin C. Kastrop,[e] Paul T. Castelein,[e] Eladio DeLeon, Jr.,[e] Steven A. Dugoni,[e] and Chun-Hsi Chung[e]

Philadelphia, Pa, Houston, Tex, Boston, Mass, Marquette, Mich, Billings, Mont, Princeton, Ill, Augusta, Ga, and South San Francisco, Calif

Posttreatment tooth movement is inevitable, but its degree depends on a variety of factors that are both iatrogenic and innate to each patient. Although various retention techniques have been developed to minimize posttreatment movement, it is unrealistic to believe that the entire dentition can be retained in all dimensions. Relapse is usually considered an adverse phenomenon, but some dimensions of posttreatment tooth movement might actually enhance occlusal function and esthetics. Favorable movement is often considered as "settling" if that aspect of the occlusion is improved over time.

In 1998, the American Board of Orthodontics (ABO) published an objective method of evaluating posttreatment results[1] by using 7 cast measurements and 1 radiographic measurement. Formerly called the Objective Grading System, the ABO now refers to this as the Cast Radiograph Evaluation (C-R Eval). This evaluation technique is used to score final casts and the panoramic radiograph produced within 12 months of debanding. The 8 scoring parameters are alignment/rotations, marginal ridges, buccolingual inclination, overjet, occlusal contacts, occlusal relationships, interproximal contacts, and root angulation. The C-R Eval was developed to increase objectivity in evaluation of treatment results; it uses a point system with precise criteria for cast and radiographic grading to score the case result in terms of finishing and detailing. Each case is scored after individual and group calibration of examiners in an effort to secure equity in grading among all examiners. Hence, 8 aspects of the finished result can be quantified at the time of final records production.

Currently, all 8 parameters of the C-R Eval are weighted equally with 1 point for each deficiency. The deficiencies are clearly defined to the accuracy of 0.5 mm. It became apparent that possibly the parameters that show improvement after treatment should be scored less heavily than those that tend to deteriorate after treatment. In other words, treatment deficiencies that will tend to relapse unfavorably should be scored more heavily than parameters that improve after debanding. The difference in scoring between improving and deteriorating parameters will be herein referred to as weighting. This information can also be helpful to the orthodontic practitioner to prognosticate which aspects of the final treatment result might settle toward the ideal and which will deteriorate over time.

The purpose of this article was to determine posttreatment tooth position changes in orthodontic cases with the intent to predict favorable vs unfavorable movement. This information is to be used to weight certain parameters of the finished result as scored by the C-R Eval. A secondary purpose of this investigation was to compare tooth movement between 2 forms of retention.

Our 2 null hypotheses were that all 7 (cast assessment) parameters of the C-R Eval have equal magnitudes of posttreatment changes, and that fixed and removable retention methods show equal amounts of posttreatment movement.

It is well established that posttreatment changes occur in orthodontic patients, but prediction of favorable vs unfavorable movement is difficult. Little et al[2] demonstrated that more deterioration of alignment occurs between 10 and 20 years after treatment than from debanding to 10 years posttreatment. Only 10% of their 100-patient sample had acceptable alignment after 10 years. Fidler et al[3] noted that posttreatment Class II Division 1 patients relapsed as much as 3.5 mm at first molar and canine relationships, 3 mm in overjet, and 4.5 mm in overbite after treatment. On a more positive note, Razdolsky et al[4] assessed 40 patients at up to 21

[a]Immediate past president of the American Board of Orthodontics (ABO).
[b]President of the ABO.
[c]President-elect of the ABO.
[d]Secretary-treasurer of the ABO.
[e]Director of the ABO.
Reprint requests to: Peter M. Greco, 834 Chestnut Street, Suite M209, Philadelphia, PA 19107; e-mail, pgrecodmd@gmail.com.
Submitted and accepted, June 2010.
Am J Orthod Dentofacial Orthop 2010;138:552-8
0889-5406/$36.00

doi:10.1016/j.ajodo.2010.06.002

months posttreatment and found continuous increases in interarch tooth contacts by 3 months, but buccolingual relationships showed minimal changes. Nett and Huang[5] used the ABO's cast radiographic evaluation (called the OGS in that article). They assessed 100 random subjects at least 10 years postretention and found that although adverse alignment changes occurred after debanding, significant improvement in marginal ridges, buccolingual relationships, occlusal contacts, and overjet occurred. Occlusal relationships improved insignificantly. None of these studies proposed a weighting factor for the C-R Eval scoring instrument.

Tooth movement during fixed and removable retention is also recognized. Although few studies compared fixed vs removable retention methods with regard to amount of relapse or settling, 1 such study was conducted by Atack et al,[6] whose 58-patient sample demonstrated no difference in mandibular incisor irregularity between bonded canine-to-canine retainers and removable (spring aligner) appliances. They found that movement occurred with either retainer design. They speculated that the movement in the fixed retention group could be related to activity in the bonded wire or wire deformation while the retainer was in place. In a study conducted by Sari et al,[7] 50 orthodontic patients were evaluated by assessment of interocclusal records produced by an impression material technique to record changes in occlusal contacts over time. Twenty-five subjects received maxillary and mandibular Hawley retainers, and 25 used maxillary and mandibular bonded anterior retainers. They also evaluated a control group of 20 subjects. The retained patients were assessed between 14 and 15 months posttreatment, and the control group was assessed at 12 months. These authors concluded that although there were insignificant changes in the control sample, both retention groups had greater numbers of occlusal contacts. More occlusal contacts were noted in the fixed retainer group than in the removable retainer group.

MATERIAL AND METHODS

One hundred twenty-six diplomates certified by the ABO in February 2006 and 2007 were contacted by form letter and requested to produce a set of casts of the cases that were successfully displayed at the ABO Clinical Examination. Sixteen diplomates responded by sending at least 1 case to the ABO central office. These diplomates were informed only that the ABO was assessing case stability and that the casts would be evaluated anonymously. They were also assured that the findings of the investigation would remain anonymous and hence would not affect their status as an ABO diplomate. The diplomates were asked to produce casts of these patients between 12 and 24 months after debanding. The diplomates were directed to trim the casts to maximum intercuspation and mail them to the ABO central office in St Louis. These casts were called settling casts. Radiographs for assessment of the root paralleling parameter on the C-R Eval were not requested.

The 2008 directors of the American Board of Orthodontics were calibrated and instructed to individually score the settling casts using the C-R Eval system. These directors will hereafter be called the examiners in this article. The cast assignments of each examiner were randomized. These scores were compared with the scores recorded for these same cases at the February 2006 and 2007 ABO Clinical Examinations.

If fixed (bonded) retention was used in either arch, this was noted. There was no other assessment of retention method.

RESULTS

A total of 50 sets of settling casts were returned to the ABO for scoring. A summary of the examiner distribution for scoring is as follows: 41 cases were scored by 2 examiners, 7 cases were scored by 3 examiners, and 2 cases were scored by 4 examiners, with an average of 2.2 examiners assessing each settling case.

Settling time was defined by the date when the final treatment casts presented in the ABO Examination Case Displays were obtained subtracted from the date that the settling casts were obtained. These times can be summarized as follows: shortest, 9 months; longest, 116 months (9 years 8 months); average, 39.6 months (3 years 4 months).

Note that 50% of the cases were scored in 35 months of settling time (2 years 11 months).

Each case was scored by 2 to 4 examiners during the investigation. For the purpose of clarity here, the original C-R Eval (cast) score will be titled the examination score (ES). This is the scoring of the case when it was originally presented at the ABO Clinical Examination in 2006 or 2007. The subsequent rescoring conducted by the ABO examiners for this investigation will be called the settling score (SS).

Determination of examiner variability and agreement

The scoring of the settling casts was randomized because each examiner scored according to the availability of the casts and time constraints. The internal agreement of the SS was evaluated by determining the variation in each examiner's range. To evaluate the interexaminer agreement on scoring, the maximum

Table I. Ranges of interexaminer settling scores

	Alignment/rotations	*Marginal ridges*	*Buccolingual inclination*	*Overjet*	*Occlusal contacts*	*Occlusal relationships*	*Interproximal contacts*	*Total*
Shortest range	0	0	0	0	0	0	0	0
Largest range	6	5	10	13	8	6	2	22
Average range across 50 cases	2.1	1.8	3.5	2.0	1.8	1.5	0.1	6.1

Table II. Intraclass correlations of the examiners

ICC	*Alignment/rotations*	*Marginal ridges*	*Buccolingual inclination*	*Overjet*	*Occlusal contacts*	*Occlusal relationships*	*Interproximal contacts*	*Total*
Agreement (ICC) for cases by 2 examiners (41 cases)	.86	.46	−.72	.64	.47	.80	.87	.69
Agreement (ICC) for cases by 3 examiners (9 cases)	.81	.89	.63	.53	.82	.85	.94	.85
Overall agreement (ICC) for all cases	.80	.57	−.28	.49	.51	.79	.87	.67

absolute difference between the SS values was calculated for each case. The lowest SS was subtracted from the highest SS, and this difference reflected the range of scores. For example, if 3 examiners gave scores of 3, 4, and 6, the score range was 6 – 3 = 3. Smaller ranges reflect closer values between examiners and, hence, less variation.

After the range was calculated for each SS, this information was summarized among the 50 cases. The shortest range was zero for all cases, indicating that the examiners reached absolute agreement on at least 1 case. For other cases, however, there were greater differences between examiners. For example, SS in the overjet parameter differed by 13 for 1 case and by 20 for another. The average ranges across 50 cases are reported in the bottom row of Table I.

The SS difference describes the degree of examiner disagreement despite calibration among examiners. This difference is sensitive to outliers and the clinical range of each parameter. Measurements that were larger in range were thus more likely to display more variation than those of smaller range.

The second method used to evaluate the internal agreement between examiners and their SS was the intraclass correlation (ICC). Since the examinees and examiners were a random sample, the 2-way random ICC assessment was chosen as the appropriate agreement measure for this study. For cases evaluated by 2 examiners, both examiners' data were used to calculate an ICC estimate. For cases with more than 2 examiners, 2 examiners' scores were randomly selected from the 3 or 4 examiners' scores. Finally, all cases were combined, and an overall agreement was estimated for each parameter. Table II gives the ICC values of the examiners.

An ICC of approximately .60 or greater would indicate sufficient agreement. Using this standard, the examiners achieved agreement on the following measures: alignment/rotations, occlusal relationships, interproximal contacts, and total. The agreement was moderate for marginal ridges, overjet, and occlusal contacts. The agreement was low for buccolingual inclination.

When the ES was compared with the SS, the average SS was used despite the varying levels of agreement because the average SS still provided a good summary of all examiners and allowed a straightforward assessment of changes over time.

Assessment of tooth position changes over time

The ES vs SS, regardless of the duration of settling time or mode of retention, was compared. The SS was obtained by averaging the scores of all examiners that reviewed the settling cases. The examiners' scores were compared by using a paired-sample *t* test as depicted in Table III.

Paired *t* test results confirmed that the scores for alignment, buccolingual inclination, and total changed significantly during settling, regardless of settling time. The other parameters had relative stability without respect to settling time.

From a graphic perspective, the ES average of 50 cases is depicted in the left bar in each graph and the SS is on the right of the Figure. Note the relatively large discrepancy in the alignment/rotation and buccolingual inclination parameters. The other parameters displayed few changes between ES and SS. The whisker extension on the solid bar graph indicates the upper boundary of measurement.

Table III. Examination score (ES) and settling score (SS) compared with paired-sample *t* tests

	ES		SS			
	Mean	*SD*	*Mean*	*SD*	*SS-ES difference*	P
Alignment/rotations	2.6	1.8	5.4	2.8	2.8	<0.01
Marginal ridges	3.4	1.7	3.0	1.6	−0.3	0.22
Buccolingual inclination	2.6	2.2	3.9	1.8	1.3	<0.01
Overjet	2.3	2.2	2.2	2.4	−0.1	0.77
Occlusal contacts	2.2	2.2	2.0	1.7	−0.2	0.66
Occlusal relationships	2.2	2.2	1.9	2.3	−0.3	0.42
Interproximal contacts	0.4	0.9	0.2	0.5	−0.2	0.28
Total	15.6	5.8	18.6	6.8	3.0	<0.01

Comparison of settling over 2-year time intervals

As previously stated, the diplomates had been asked to submit settling casts produced between 12 and 24 months after debanding, but many casts were received that were produced more than 2 years after debanding. These data were used in the following analysis to determine settling activity over defined time intervals.

ES and SS classified by settling time were stratified into periods as listed in Table IV. Standard deviations are given in parentheses.

The average changes (defined as the SS minus the ES) are reported in Table V by these time periods: less than 2 years, 2 to 4 years, and more than 4 years. The 3 time periods were compared by using 1-way analysis of variance (ANOVA). *P* values less than .05 indicated a significant correlation between tooth movement and settling time.

One-way ANOVA tests were used to test the difference between SS and ES across time periods. Results indicated that significant movement occurred for occlusal contacts (interarch occlusal contacts, $P = .046$). SS values in this parameter were higher than ES values for cases evaluated within 4 years but lower for cases evaluated beyond 4 years. This indicates that most occlusal contact improvement occurs in the first 4 years in retention. No significance was found in the other parameters.

Assessment of settling comparing removable retention vs fixed retention

A comparison of settling characteristics of cases with fixed vs removable retention was also conducted. Twenty-five of the 50 cases were retained solely with removable retention. Eighteen used bonded retention comprising palatal wires spanning the maxillary lateral incisors and mandibular wires spanning the canines (upper 2-2/lower 3-3 retention). Seven used mixed methods and were thus excluded from this portion of the study. This left a total fixed retainer sample of 18.

The change in settling was then compared for cases with removable retention and fixed retention by using a 2-group independent *t* test. These results are listed in Table VI. A significant difference was found with buccolingual inclination ($P = 0.025$). For cases with removable retention, the SS was higher than the ES by 2.08, but only higher by 0.53 for cases with fixed retainers. This indicated that SS values are similar in all parameters except in the buccolingual inclination parameter, where more movement was noted in the removable retention group. This might be an unjustified conclusion given the unreliability of measurement of the buccolingual inclination parameter as previously mentioned.

DISCUSSION

These results demonstrate that some parameters of a completed case remain stable over time, but others do not. This information might be useful to the clinician in planning retention and case finishing. In the assessment of the finishing and detailing of completed cases in an examination arena, it is logical that parameters that will settle favorably should be penalized less severely for the deficiency than those that will deteriorate over time.

In this investigation, alignment/rotations and buccolingual inclination as assessed by the criteria of the C-R Eval (cast analysis) tended to deteriorate over time, whereas other parameters did not. Thus, the final alignment of all teeth and the buccolingual inclination of the posterior quadrants cannot be expected to improve with time, and the quality of these parameters will never be more precisely detailed than at debanding. These results support the need to increase the scoring penalty for alignment deficiencies noted at approximately the time of debanding, since this parameter is not expected to improve over time. Interexaminer reliability in the measurement of buccolingual inclination is less than ideal, however, indicating low reliability in the scoring of this parameter. Weighting of the buccolingual parameter might not be justified until an improved scoring technique is developed.

Given the 1.1 point mean settling change in alignment, a 0.5 point weighting factor in the scoring of this parameter should be considered. This value was suggested because of its ease in computation of the total score. Such weighting would mean that the C-R Eval scoring total for the alignment parameter would be increased by a factor of 1.5 to reflect the tendency of this parameter to deteriorate, rather than improve,

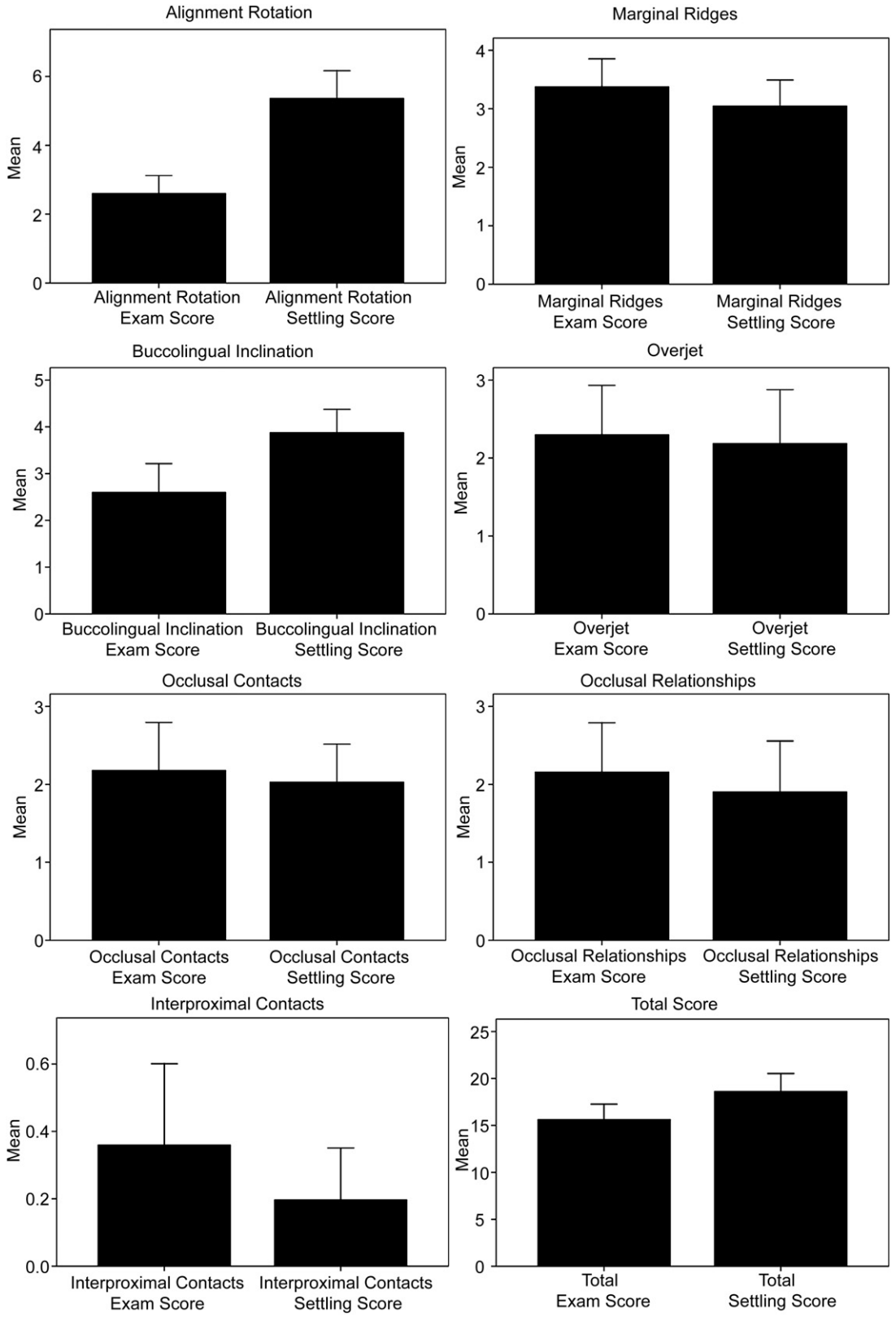

Fig. Graphic comparison of examination scores vs. settling scores for the entire 50-case sample.

Table IV. Examination scores and settling scores correlated with retention time (mean [SD])

	Less than 2 years (n = 13)		*2-4 years (n = 23)*		*4+ years (n = 14)*	
	ES	*SS*	*ES*	*SS*	*ES*	*SS*
Alignment/rotations	2.62 (1.98)	4.96 (2.80)	2.91 (1.88)	5.87 (2.85)	2.07 (1.64)	4.91 (2.95)
Marginal ridges	3.00 (1.58)	2.77 (1.30)	3.48 (1.93)	3.08 (1.72)	3.57 (1.34)	3.25 (1.65)
Buccolingual inclination	2.69 (2.72)	4.12 (1.65)	2.70 (2.14)	3.77 (1.61)	2.36 (1.74)	3.84 (2.15)
Overjet	2.15 (2.23)	1.94 (2.43)	2.52 (2.33)	1.83 (1.69)	2.07 (2.20)	3.02 (3.32)
Occlusal contacts	1.62 (1.98)	2.01 (1.85)	2.04 (1.69)	2.38 (1.85)	2.93 (2.89)	1.46 (1.26)
Occlusal relationships	1.69 (2.14)	1.65 (1.66)	2.43 (2.50)	2.32 (2.78)	2.14 (1.88)	1.45 (1.97)
Interproximal contacts	0.15 (0.55)	0.54 (0.90)	0.35 (0.78)	0.12 (0.35)	0.57 (1.16)	0.00 (0.00)
Total	13.92 (6.18)	17.99 (7.84)	16.43 (5.13)	19.37 (6.37)	15.86 (6.49)	17.93 (6.71)

Table V. Average changes (mean [SD]) over time (SS – ES)

	Less than 2 years (n = 13)	*2-4 years (n = 23)*	*4+ years (n = 14)*	P
Alignment/rotations	2.35 (2.06)	2.96 (2.98)	2.84 (2.61)	0.800
Marginal ridges	−0.23 (2.13)	−0.40 (1.96)	−0.32 (1.75)	0.969
Buccolingual inclination	1.42 (1.79)	1.07 (1.73)	1.48 (3.06)	0.828
Overjet	−0.22 (2.50)	−0.70 (2.54)	0.95 (3.04)	0.203
Occlusal contacts	0.40 (1.51)	0.34 (2.56)	−1.46 (2.30)	0.046
Occlusal relationships	−0.04 (2.34)	−0.12 (2.18)	−0.70 (2.36)	0.696
Interproximal contacts	0.38 (1.14)	−0.22 (0.85)	−0.57 (1.16)	0.057
Total	4.06 (5.82)	2.93 (7.28)	2.07 (4.86)	0.716

Table VI. Comparison of removable vs fixed retention (using independent *t* test; mean [SD])

	Removable retention (n = 25)			*Fixed retainers (n = 18)*			
	ES	*SS*	*SS-ES difference*	*ES*	*SS*	*SS-ES difference*	P
Alignment/rotations	2.52 (1.90)	5.65 (3.13)	3.13 (2.80)	2.83 (2.04)	4.78 (1.97)	1.94 (1.95)	0.129
Marginal ridges	3.08 (1.32)	3.12 (1.42)	0.04 (1.58)	3.39 (2.03)	3.15 (1.46)	−0.24 (1.91)	0.601
Buccolingual inclination	2.24 (1.83)	4.32 (1.70)	2.08 (1.94)	3.00 (2.52)	3.53 (1.71)	0.53 (2.40)	0.024
Overjet	1.88 (2.20)	2.03 (2.39)	0.15 (2.34)	2.61 (2.23)	2.40 (2.95)	−0.21 (2.99)	0.661
Occlusal contacts	1.80 (2.02)	1.69 (1.31)	−0.11 (2.26)	3.17 (2.36)	2.51 (2.12)	−0.66 (2.77)	0.482
Occlusal relationships	2.92 (2.53)	2.34 (2.31)	−0.58 (2.34)	1.56 (1.65)	1.59 (2.39)	0.04 (2.52)	0.414
Interproximal contacts	0.16 (0.47)	0.19 (0.58)	0.03 (0.79)	0.67 (1.24)	0.11 (0.47)	−0.56 (1.38)	0.084
Total	14.60 (6.08)	19.34 (6.85)	4.74 (5.27)	17.22 (5.70)	18.12 (7.31)	0.90 (7.28)	0.051

over time. Maintenance of present scoring values of all other parameters is appropriate.

In the analysis of the settling times, half of the cases had settling times within 39 months, and all but 3 were within 83 months. The longest were 116, 104, and 101 months. When the statistical analyses were repeated after eliminating the cases with settling times of 101 months or more, the results were similar. The only difference occurred in the buccolingual parameter because the elimination of these 3 values then indicated that the difference in buccolingual scores was insignificant.

In contrast to the results of Nett et al,[5] parameters other than alignment/rotations and buccolingual inclination in this study did not improve over time, but remained relatively static. Their subject group, however, was at least 10 years postretention; therefore, an extended time after debanding might be necessary to effect additional tooth movement as was observed in their study.

When considering the timing of tooth movement, data from this investigation indicated that occlusal settling (increase in interarch contacts) had ceased by 4 years after debanding.

The settling characteristics in comparison of fixed vs removable strategies used in the cases evaluated in this investigation showed similar difference in settling

between retention modalities except for buccolingual relationships. Our results paralleled the findings of Atack et al,[6] who used the irregularity index to document significant mandibular anterior movement with either Hawley retainers or multi-stranded bonded lingual wires. Those investigators found that movement occurs regardless of the retention method. The conclusions of Sari et al[7] via the use of interocclusal registration records also confirmed the increase of interarch contacts with either method of retention. More posterior contacts resulted in patients wearing anterior fixed retainers vs bimaxillary removable retention. The observation period in this study did not exceed 15 months, however, and most posterior settling in our study occurred up to 4 years after debanding.

CONCLUSIONS

1. The ABO C-R Eval demonstrates adequate interexaminer reliability in the alignment/rotations, occlusal relationships, and interproximal contacts parameters. Moderate interexaminer reliability was observed in marginal ridges, overjet, and occlusal contacts. Weak interexaminer reliability in buccolingual inclination was noted.
2. Whereas alignment deteriorates during settling, all other parameters do not.
3. Since alignment of the dentition deteriorates over time, it is appropriate to consider weighting this parameter for examination purposes because this aspect of the finished case will not improve during retention.
4. No conclusions regarding buccolingual inclination changes can be made given the questionable interexaminer reliability of this parameter.
5. When comparing case settling, the most significant improvement of occlusal contacts occurs before 4 years posttreatment.
6. There is little difference in the amount of change in occlusal contacts in the comparison of fixed vs removable retention methods.

We thank these ABO Diplomates who submitted their cases for assessment in this study: Drs John N. Archer, F. Richard Beckwith, Steven W. Covino, Frank H. Crist, David C. Gehring, Robert F. Girgis, Tom E. Hartsock, Kurt W. Hoffman, David M. Konys, Chris Mattingly, D. Spencer Pope, Viney P. Saini, Lisa L. Sanfelippo, Gilpatrick Schmidtke, Josephine C. Weeden, and Laurene A. Wolf.

REFERENCES

1. Casko JS, Kokich V, Vaden J, Damone J, James RD, Cangialosi TJ, et al. Objective grading system for dental casts and panoramic radiographs. Am J Orthod Dentofacial Orthop 1998;114:589-99.
2. Little RM, Reidel R, Årtun J. An evaluation of changes in mandibular anterior alignment from 10 to 20 years postretention. Am J Orthod Dentofacial Orthop 1988;93:423-8.
3. Fidler BC, Årtun J, Joondeph DR, Little RM. Long-term stability of Angle Class II, division 1 malocclusions with successful occlusal results at the end of active treatment. Am J Orthod Dentofacial Orthop 1995;107:276-85.
4. Razdolsky Y, Sadowsky C, BeGole E. Occlusal contacts following orthodontic treatment: a follow-up study. Angle Orthod 1989;59: 181-5.
5. Nett BC, Huang GJ. Long-term posttreatment changes measured by the American Board of Orthodontics objective grading system. Am J Orthod Dentofacial Orthop 2005;127:444-50.
6. Atack N, Harradine N, Sandy JR, Ireland AJ. Which way forward? Fixed or removable retainers. Angle Orthod 2007;77: 954-9.
7. Sari Z, Uysal T, Basciftci FA, Inan O. Occlusal contact changes with removable and bonded retainers in a 1-year retention period. Angle Orthod 2009;79:867-72.