Accepted: 3 June 2017

DOI: 10.1111/ocr.12195

ORIGINAL ARTICLE

WILEY Orthodontics & Craniofacial Research

Validity of the American Board of Orthodontics Discrepancy Index and the Peer Assessment Rating Index for comprehensive evaluation of malocclusion severity

S. Liu[1] | H. Oh[2] | D. W. Chambers[3] | S. Baumrind[4] | T. Xu[5]

[1]First Clinical Division, Peking University School and Hospital of Stomatology, Beijing, China

[2]Department of Orthodontics, University of the Pacific, Arthur A. Dugoni School of Dentistry, San Francisco, CA, USA

[3]Department of Dental Practice, University of the Pacific, Arthur A. Dugoni School of Dentistry, San Francisco, CA, USA

[4]Department of Orthodontics, Craniofacial Research Instrumentation Laboratory, University of the Pacific, Arthur A. Dugoni School of Dentistry, San Francisco, CA, USA

[5]Department of Orthodontics, Peking University School and Hospital of Stomatology, Beijing, China

Correspondence
T. Xu, Department of Orthodontics, Peking University, School and Hospital of Stomatology, Beijing, China.
Email: tmxuortho@163.com

Funding information
Specific Research Project of Health Pro Bono Sector, Ministry of Health, China, Grant/Award Number: 200802056; International Science and Technology Cooperation Program of China, Award Number: 2014DFA31800

Structured Abstract

Objectives: To assess the validity of the American Board of Orthodontics Discrepancy Index (ABO-DI) and Peer Assessment Rating (PAR) Index in evaluating malocclusion severity in Chinese orthodontic patients.

Setting and Sample Population: A stratified random sample of 120 orthodontic patients based on Angle classification was collected from six university orthodontic centres.

Material and Methods: Sixty-nine orthodontists rated malocclusion severity on a five-point scale by assessing a full set of pre-treatment records for each case and listed reasons for their decision. Their judgement was then compared with ABO-DI and PAR scores determined by three calibrated examiners.

Results: Excellent interexaminer reliability of clinician judgement, ABO-DI and PAR index was demonstrated by the Intraclass Correlation Coefficient (rho= 0.995, 0.990 and 0.964, respectively). Both the ABO-DI and US-PAR index showed good correlation with clinician judgement (r=.700 and r=.707, respectively). There was variability among the different Angle classifications: the ABO-DI showed the highest correlation with clinician judgement in Class II patients (r=.780), whereas the US-PAR index showed the highest correlation with clinician judgement in Class III patients (r=.710). Both indices demonstrated the lowest correlations with clinician judgement in Class I patients.

Conclusion: With strong interexaminer agreement, the panel consensus was used for validating the ABO-DI and US-PAR index for malocclusion severity. Overall, the ABO-DI and US-PAR index were reliable for measuring malocclusion severity with significantly variable weightings for different Angle classifications. Further modification of the indices for different Angle classification may be indicated.

KEYWORDS
Discrepancy Index, malocclusion severity, Peer Assessment Rating Index, validity

1 | INTRODUCTION

An index is a measure used to represent a complex judgement. Indices should omit minor details, be easy to use, have high levels of interexaminer reliability, and correlate with the interests of practitioners and scholars. Necessarily, indices explain somewhat less than the total amount of variance in clinical judgements, but the best indices explain more variance.

Over the years, several occlusal indices have been developed in an effort to provide a more objective assessment of malocclusion

 © 2017 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd

severity.[1-4] These indices have their limitations and advantages. One of the most widely used occlusal indices is the Peer Assessment Rating (PAR) index.[2,4] The PAR index includes five components (alignment, buccal occlusion, overjet, overbite/open bite and centreline) scored from dental casts and provides a single summary score for all the occlusal anomalies.[2] Weightings have been derived for individual components from validation studies that use panel assessment to reflect the current orthodontic community's opinion in many countries. Among these weightings, the UK- and US-weighting systems are most widely used.[2,5] The PAR index has good reliability and validity,[2,5] but it does not evaluate many aspects of a malocclusion as it uses only study casts.

On the other hand, the American Board of Orthodontics Discrepancy Index (ABO-DI), which was introduced in 1998 by the American Board of Orthodontics, summarizes the clinical features of a patient's condition using pre-treatment measurements on study casts and cephalometric and panoramic radiographs.[3] The ABO-DI scores 12 target disorders: overjet, overbite, anterior open bite, lateral open bite, crowding, occlusal relationship, lingual posterior cross-bite, buccal posterior cross-bite, ANB, SN-MP, L1-MP and other components. Although the ABO-DI has been widely used to assess malocclusion complexity, few studies have investigated its relationship with clinician judgement and other indices.[6]

Using a strong set of indices will promote professional standardization, allowing for comparisons to be made nationally and internationally. Ultimately, a valid index of malocclusion severity would not only benefit practitioners by providing executable guidelines in daily practice but would also contribute to building a reference standard for the orthodontic community. However, it is necessary to test the validity of the indices against clinician judgement within the population being studied.

The purposes of this study were to assess the validity of both the ABO-DI and PAR index compared to clinician judgement and to identify similarities and differences between the ABO-DI and PAR index in evaluating malocclusion severity in Chinese orthodontic patients.

2 | MATERIAL AND METHODS

The present study was designed as a multicentre, retrospective study. Six orthodontic treatment centres located in various parts of China participated in this study: Peking University School of Stomatology (Centre 1), West China College of Stomatology at Sichuan University (Centre 2), School of Stomatology at the Fourth Military Medical University (Centre 3), Beijing Stomatological Hospital and School of Stomatology at the Capital Medical University (Centre 4), Stomatological Hospital of Nanjing Medical University (Centre 5) and Hospital of Stomatology at Wuhan University (Centre 6) (Figure 1).

To form a representative patient sample, each centre collected complete medical records for at least 250 patients who had treatment completed between July 2005 and September 2008. A stratified random sampling method based on Angle classification documented in the patient chart was used. From a combined total of 2383 records, 21 were drawn from each centre to create a sample of 126 subjects that consisted of equal numbers of Angle Class I, Class II and Class III subjects (Figure 1).

Patient records included study casts, lateral cephalometric radiographs, panoramic radiographs, facial photographs (front, lateral and front smile views) and treatment charts. Six subjects were excluded before data acquisition due to their study casts being damaged during

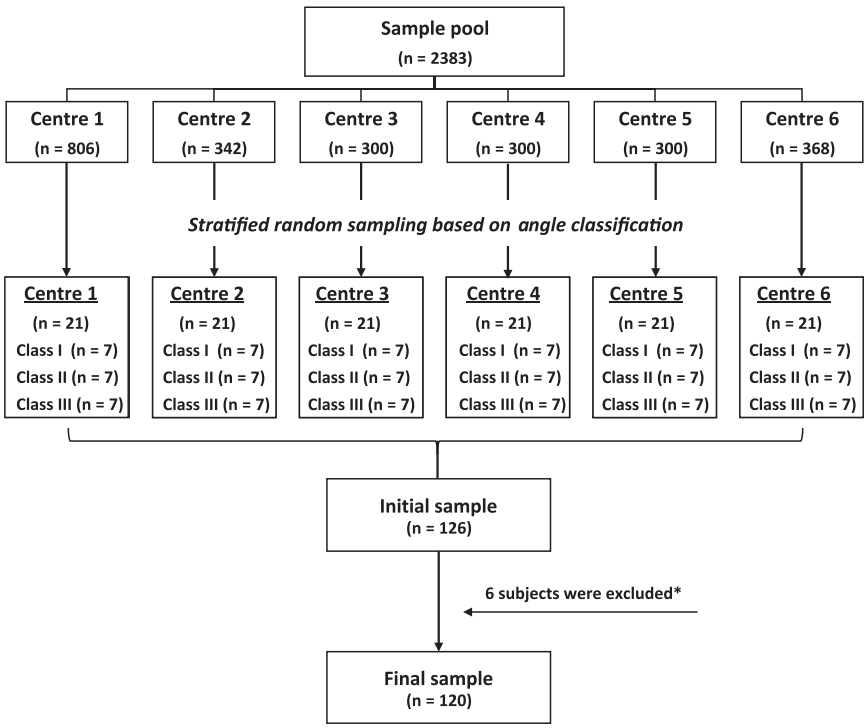

FIGURE 1 Schematic sampling process. *Six study casts were accidently broken during the evaluation session and could not be restored in time

the clinician evaluation session. The final sample consisted of 120 subjects, as described in Figure 1.

A panel of 72 judges, composed of 12 judges from each centre, was recruited initially. Inclusion criteria for judge selection were: (i) more than 8 years of clinical experience in orthodontics; (ii) an M.S. or Ph.D. degree in orthodontics; (iii) an academic rank of associate professor or above. Three judges were dropped due to schedule conflicts that prohibited them from participating in the clinician evaluation session that was held at a national meeting, leaving a total of 69 judges. Each of the 120 cases in the study was examined by the 69 judges.

Following protocol, judges were presented with anonymous cases in nine groups of 14 records each. Breaks were provided between sets of cases. Given a full set of diagnostic records, each judge rated the malocclusion severity of each patient using a five-point rating scale (1—mild, 2—mildly moderate, 3—moderate, 4—severely moderate, 5—severe) and listed at least three reasons that contributed to their assessment of malocclusion severity. The judges were verbally instructed not to consider any factors related to treatment time, methods, costs or the compliance of patients.

Indices (ABO-DI and PAR index) were calculated from measurements made on study casts and cephalometric radiographs by three-second-year orthodontic residents. Preliminary calibration sessions were carried out using ten randomly selected cases. Each examiner individually measured each case three times within a five-day interval. The intra- and interexaminers reliability was tested with Intraclass Correlation Coefficient (ICC). The categories with an ICC value of less than 0.75 were discussed and measured again. After repeating three calibration sessions, there was no category with an ICC value of less than 0.75. After 4 weeks of calibration, each examiner measured all patients in the final sample. Each component of the ABO-DI and PAR index was recorded separately and averaged between the three examiners. Then, weightings were added to the raw PAR index scores in adherence to the United Kingdom (UK) weighting system and the United States (US) weighting system.

TABLE 1 Sample demographic information

		Sex		Age	
Angle classification	Number	Male	Female	Mean	SD
Class I	38	11	27	16.3	6.5
Class II	40	12	28	18.4	7.7
Class III	42	10	32	16.6	4.3
Total	120	33	87	17.1	6.3

SD, Standard Deviation.

2.1 | Statistical analysis

Sample sizes and normality of distributions were inspected to ensure that parametric statistical analysis was appropriate. Descriptive statistics, including proportions, means and standard deviations, and bivariate correlations were calculated. Interexaminer reliability among judges and residents determining the indices was measured using Cronbach's generalizability method[7] which, in this case, gives an unbiased approximation of the Intraclass Correlation Coefficient (ICC). Tests for differences between means were performed on averages across patient groups and judges using ANOVA with post hoc Scheffe's method. The accuracy of this approach was verified by repeated measures ANOVA.

3 | RESULTS

Excellent interexaminer reliability among the 69 orthodontists rating overall severity of malocclusions was demonstrated. The mean ICC for clinician judgement was 0.995 with CI_{95} of 0.075 ($P<.001$). Interexaminer reliability of the ABO-DI and PAR index measured by the three residents was also excellent (0.990 and 0.964, respectively, $P<.001$).

The sample demographic information is shown in Table 1. There were no statistically significant differences between Angle classifications with regard to sex and age. The panel of 69 orthodontists was comprised of 35 male and 34 female orthodontists. The mean age was 45.2±7.3 years, ranging from 35 to 72 years. The clinicians had an average of 19±7.2 years of practice experience, ranging from 8 to 49 years. There were no statistically significant differences in clinician judgement related to clinician's sex, age, years of experiences or centres they practiced at.

The validity of both the ABO-DI and US-PAR index was examined by correlating both scores with clinician judgement rating, which was defined by the average of the 69 judges' assessments. At first, the bivariate correlations between clinician judgement and the raw PAR index, the PAR index with UK weighting[2] and the PAR index with US weightings (US-PAR)[5] were performed, with the US-PAR index revealing the highest correlation ($r=.431$, $r=.646$ and $r=.707$, respectively, $P<.001$). Based on that result, only the US-PAR index was retained for further analysis. Table 2 shows correlations between clinician judgement and the ABO-DI and US-PAR index. A moderate correlation was found between the ABO-DI and the US-PAR index ($r=.549$, $P<.001$). When ABO-DI was calculated using only measurements on the study

TABLE 2 Correlations between Clinician judgement, Discrepancy Index and US-PAR index

	Discrepancy Index	US-PAR index
Clinician judgement on the whole sample (n=120)	.700*	.707*
Clinician judgement on Class I sample (n=38)	.518*	.588*
Clinician judgement on Class II sample (n=40)	.780*	.591*
Clinician judgement on Class III sample (n=42)	.654*	.710*

*$P< .001$.

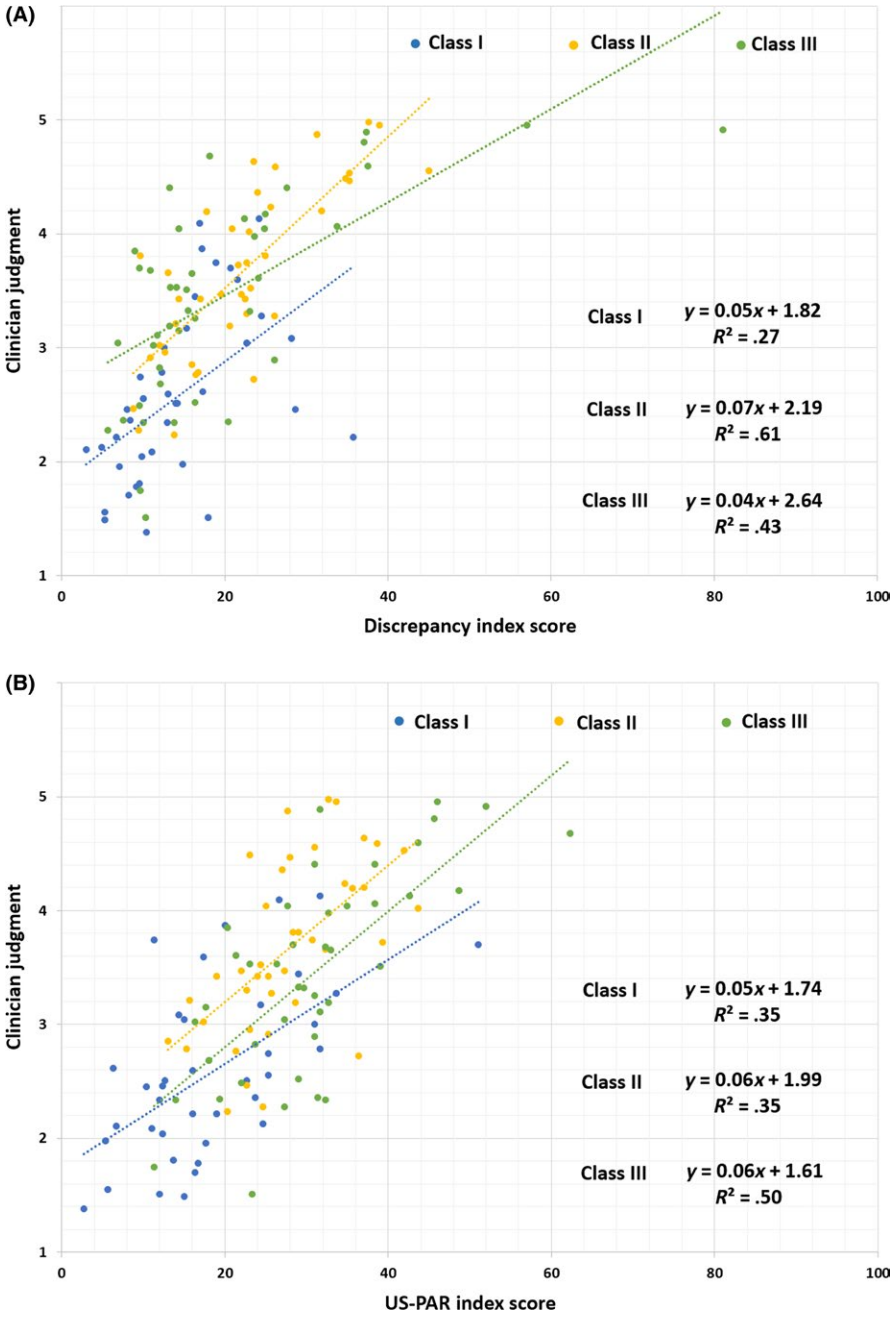

FIGURE 2 Scatterplots between clinician judgements (mean subjective scores by the panel of 69 orthodontists) and the Discrepancy Index (A) and US-PAR index (B) [Colour figure can be viewed at wileyonlinelibrary.com]

casts (excluding cephalometric measurements), a slightly higher correlation was found between the two indices (r=.599, P<.001).

Both the ABO-DI and the US-PAR index scores showed the lowest correlations with clinician judgement in Class I patients (Table 2, Figure 2). The highest correlation was found between clinician judgement and ABO-DI in Class II patients (r=.780). And a relatively high correlation between US-PAR index and clinician judgement was found in Class III patients (r=.710).

Table 3 shows the mean values for clinician judgement, ABO-DI and US-PAR index scores by Angle classification. Clinician judgement and ABO-DI showed the same pattern, with the highest mean values in the Class II group and the lowest mean values in the Class I group. In contrast, the US-PAR index showed the highest mean scores in the Class III group. For all three methods, a post hoc Duncan multiple range tests demonstrated that Class I patients had a lower score than that of Class II or Class III patients, but there was no statistically significant difference between Class II and Class III patients.

There was no statistically significant difference between male and females patients for all three methods. However, patient age was mildly associated with clinician judgement (r=.339, P<.001) and ABO-DI (r=.303, P<.001), but no correlation was found between patient age and the US-PAR index (r=.063).

To explore the reasons for clinician judgement made on the three Angle classification groups, the frequency of reasons listed by judges during their evaluation was examined after all reasons were computerized and assigned categorization.[8] The top five reasons that influenced

TABLE 3 Means and Standard Deviation (SD) of Clinician judgement, ABO-Discrepancy Index and US-PAR index by Angle classification

	Class I (n=38)		Class II (n=40)		Class III (n=42)				
	Mean	SD	Mean	SD	Mean	SD	Differences	F	*P*
Clinician judgement	2.58	0.76	3.67	0.75	3.45	0.89	I<II, I<III	19.76	<.001
ABO-Discrepancy Index	14.38	7.56	22.15	8.80	19.73	13.94	I<II, I<III	5.37	.01
US-PAR index	18.35	10.44	27.76	8.57	30.37	10.04	I<II, I<III	19.02	<.001

clinician judgement for each Angle classification group are illustrated in Figure 3. They were dental crowding, skeletal sagittal pattern, soft tissue profile, molar relationship, overbite and overjet. Interestingly, overjet was included in the top 5 reason for the Class II group, but not for the Class I and Class III groups.

4 | DISCUSSION

The panel of orthodontists in this study provided an excellent interexaminer reliability that was comparable to previous related reports.[2,5,9] In addition, there were no statistically significant differences in clinician judgement of malocclusion severity that were related to the clinician's sex, age, number of years of specialty experience or centres where the clinicians practiced. The excellent interexaminer reliability in the present study suggests that clinician judgements can be used as a reliable reference for validating the indices.

Correlation between clinician judgement and PAR index in the present study was slightly lower compared to previous reports by Richmond et al. and DeGuzman et al.[2,5] (r=.85 and r=.83, respectively). This may be explained by the fact that study casts were the only records used for clinician judgement rating in those previous studies[2,5] whereas a full set of records that included lateral cephalometric radiographs, facial photographs and study casts was used for clinician judgement in the present study. The previous literature has shown that cephalometric radiographic information in addition to study casts can influence the clinician's decision on treatment planning and increase the variability of their decisions for Class II patients.[10,11] This finding was reflected in the clinicians' reasons for determining malocclusion severity that was collected in this study; the skeletal sagittal pattern was one of the most frequently listed primary reasons in influencing a clinician's decision when evaluating the severity of malocclusion in Class II and Class III patients (Figure 3).

As cephalometric measurements are utilized in the ABO-DI, one would assume that the ABO-DI better correlates with clinician judgement than the US-PAR index that is solely based on occlusal traits. However, this only held true for Class II patients. These data suggest, though not conclusively, that the ABO-DI showed a higher correlation with clinician judgement for Class II patients than the US-PAR index (P=.058). A lower correlation of the ABO-DI than the US-PAR index with clinician judgement in the Angle Class III group was an unexpected finding that requires further investigation.

Another interesting finding was the relatively low correlation between both the ABO-DI and US-PAR index and clinician judgement for the Angle Class I group as compared to the Angle Class II and III groups. As supported by Pae et al.'s study, clinicians judged bimaxillary protrusion with a protrusive facial profile in Angle Class I patients as a severe form of malocclusion.[12] This suggests that variations in soft tissue profile may contribute to relatively lower correlations between clinician judgement and indices for Class I patients than for Class II and III patients. The reasons for clinician judgement of severity support these findings. In addition, although arch crowding was listed as the most influential reason for judging Class I severity in this study, which was also supported by Konstantonis et al.,[13] it may not be well reflected by the indices as the US-PAR index only scores the upper anterior alignment and the ABO-DI only includes the most crowded arch.

There are some reports that consider age or growth potential when assessing the malocclusion severity.[14] The present study indicates that clinician judgement loosely correlated with patient age. The ABO-DI also showed a similar correlation with age, but this was not the case for the US-PAR index.

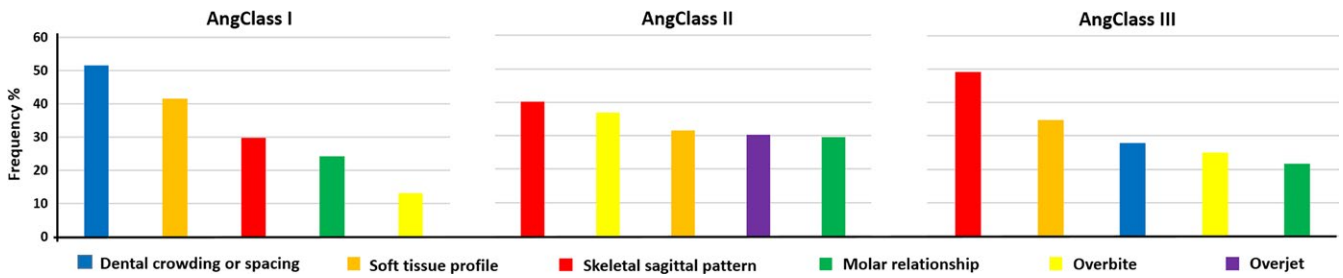

FIGURE 3 Illustration of top five reasons that influenced clinician judgement of malocclusion severity ratings for each Angle classification group [Colour figure can be viewed at wileyonlinelibrary.com]

Useful information was obtained from this study by comparing the two most widely used indices with clinician judgement in measuring the severity of malocclusion in a Chinese orthodontic patient population. It seems that both the ABO-DI and the US-PAR index can serve as reasonable approximations of a clinician's overall judgement of malocclusion severity (r=.700 and r=.707, respectively), but they can only explain about 50% of the variability in clinician judgement. Moreover, varying degrees of strength of associations between the indices and clinician judgement for different Angle classifications indicate that the current indices are not uniformly valid in representing clinician judgement for different Angle classifications.

These results were derived from a panel of orthodontists practicing in China. It is possible that the perception of the severity of malocclusion may vary between different racial backgrounds and geographical locations, and thus, further validation exercises and modification of indices may be indicated.

5 | CONCLUSIONS

Excellent interexaminer agreement among orthodontists in evaluating malocclusion severity was demonstrated regardless of a clinician's sex, age, years of experience and location of his/her centre. Overall, the ABO-DI and US-PAR index were reliable and similarly correlated with clinician judgement in measuring malocclusion severity. However, varying degrees of the strength of associations between the indices and clinician judgement for different Angle classifications were found.

ACKNOWLEDGEMENTS

We express gratitude to the 69 orthodontists who served as judges and the three residents, Xuanrong Weng, Qing Chen and Chong Ren who worked as examiners.

REFERENCES

1. Summers CJ. The occlusal index: a system for identifying and scoring occlusal disorders. *Am J Orthod*. 1971;59:552-566.
2. Richmond S, Shaw WC, O'brien KD, et al. The development of the PAR Index (Peer Assessment Rating): reliability and validity. *Eur J Orthod*. 1992;14:125-139.
3. Cangialosi TJ, Riolo ML, Owens SE, et al. The ABO discrepancy index: a measure of case complexity. *Am J Orthod Dentofacial Orthop*. 2004;125:270-278.
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;14:180-187.
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;107:172-176.
6. Kenneth Britt Reagin. The American Board of Orthodontics discrepancy index: an evaluation of its validity [Abstract]. *Am J Orthod Dentofacial Orthop*. 2006;130:806.
7. Cronbach J, Gleser GC, Nanda H, Rajaratnam N. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: John Wiley & Sons; 1972.
8. Baumrind S, Korn EL, Boyd RL, Maxwell R. The decision to extract: part II. Analysis of clinicians' stated reasons for extraction. *Am J Orthod Dentofacial Orthop*. 1996;109:393-402.
9. 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;122:463-469.
10. Nijkamp PG, Habets LL, Aartman IH, Zentner A. The influence of cephalometrics on orthodontic treatment planning. *Eur J Orthod*. 2008;30:630-635.
11. Silling G, Rauch MA, Pentel L, Garfinkel L, Halberstadt G. The significance of cephalometrics in treatment planning. *Angle Orthod*. 1979;49:259-262.
12. Pae EK, McKenna GA, Sheehan TJ, Garcia R, Kuhlberg A, Nanda R. Role of lateral cephalograms in assessing severity and difficulty of orthodontic cases. *Am J Orthod Dentofacial Orthop*. 2001;120:254-262.
13. Konstantonis D, Anthopoulou C, Makou M. Extraction decision and identification of treatment predictors in Class I. *Prog Orthod*. 2013;14:47.
14. Schafer SM, Maupome G, Eckert GJ, Roberts WE. Discrepancy index relative to age, sex, and the probability of completing treatment by one resident in a 2-year graduate orthodontics program. *Am J Orthod Dentofacial Orthop*. 2011;139:70-73.

How to cite this article: Liu S, Oh H, Chambers DW, Baumrind S, Xu T. Validity of the American Board of Orthodontics Discrepancy Index and the Peer Assessment Rating Index for comprehensive evaluation of malocclusion severity. *Orthod Craniofac Res*. 2017;20:140–145. https://doi.org/10.1111/ocr.12195