Validity and Reliability of the Photographic Open Bite Severity Index (POSI)

Sherry X. Wan, DMD, MSE

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Dentistry

University of Washington

2023

Committee:

Burcu Bayirli

Greg Huang

Don Joondeph

Lloyd Mancl

Program Authorized to Offer Degree:

Orthodontics

University of Washington

ABSTRACT

Validity and Reliability of the Photographic Open Bite Severity Index (POSI)

Sherry Wan

Chair of the Supervisory Committee:

Burcu Bayirli

Department of Orthodontics

Introduction: This study aimed to assess the validity and reliability of the Photographic Open Bite Severity Index (POSI), a recently developed measure of anterior open bite (AOB) severity. In addition, a formula to combine the POSI score and the vertical millimetric measure of AOB into the Combined Open Bite Severity Index (COSI) was developed to serve as a more comprehensive quantitative assessment of AOB severity.

Methods: An experienced panel of 5 orthodontists viewed 84 intraoral frontal photos of adult AOB patients at two times to provide their subjective clinical evaluations of AOB severity and treatment difficulty, and a third time, to assign POSI scores. Inter- and intra-examiner reliability of clinical evaluation scores were assessed using the intraclass correlation coefficient and examiner-assigned POSI scores were compared to gold standard POSI scores for accuracy with the weighted kappa. Digitally-designed 3D printed AOB models were photographed from above and below the occlusal plane to assess the effects of photo angulation on POSI reliability. Correlations between examiners' subjective clinical evaluations and POSI score were evaluated by Spearman rank correlation for POSI validity. Linear regression was used for the creation of the COSI formula.

Results: The panel's subjective clinical evaluations of AOB severity and treatment difficulty showed excellent intra-examiner reliability (ICCs between 0.78 and 0.97) and good inter-examiner reliability (ICC, 0.76 and 0.67, respectively). The examiner-assigned POSI scores were reliable (weighted kappas between 0.94 to 0.97) and accurate when compared to gold-standard POSI scores (weighted kappas between 0.90 and 0.96). Photographs of models obtained from ≤10 degrees above or below the occlusal plane resulted in the same POSI score as that of the un-angulated models. Correlations between gold standard POSI scores and both examiner-assigned AOB severity and treatment difficulty were moderate (Spearman rank correlation, 0.71 and 0.66, respectively). The combination of POSI and estimated clinical OB resulted in the highest correlations with AOB severity and treatment difficulty (correlation: 0.80 and 0.76, respectively) and they were used to create the COSI.

Conclusion: The Photographic Open Bite Severity Index (POSI) is a valid and reliable measure of severity and treatment difficulty of anterior open bite (AOB), when scored on an intraoral frontal photograph captured from ≤10 degrees above or below the occlusal plane. While the correlations between POSI alone, estimated clinical OB alone, POSI + estimated clinical OB, and POSI + ceph-measured OB with AOB severity and treatment difficulty were all strong, the combination of POSI + estimated clinical OB (the COSI) produced the highest correlation. The POSI and the COSI can both be utilized as a standardized outcome measure in clinical research or as a measure for AOB in determining treatment need in public health and insurance domains.

ACKNOWLEDGEMENTS

I would like to thank the University of Washington Department of Orthodontics and the University of Washington Alumni Association for their support in this rewarding academic pursuit. Thank you to my thesis committee members, Drs. Burcu Bayirli, Greg Huang, Don Joondeph, Lloyd Mancl, and Geoff Greenlee for their time, guidance, and contributions. In addition, I could not have completed this work without the five faculty members who volunteered to participate in my research project as panel members, and for that I am immensely grateful. I would also like to express my thanks and appreciation for my fiancé, friends, and family for their encouragement and support. Finally, thank you to Teddy, Monty, and Riley for being the sweetest and most loving canine companions.

TABLE OF CONTENTS

INTRODUCTION

Anterior open bite (AOB) is defined as a lack of vertical overlap of the anterior dentition in maximum intercuspation (Fields et al. 2019). The prevalence of this condition is reported to be between 0.6% to 16.5% in the United States, varying by age and race (Kelly et al. 1973; Proffit et al. 1998). While the prevalence of this malocclusion may appear relatively low, it is a condition that results in significant esthetic and functional distress for patients, including problems with enunciation and difficulty incising food during mastication (Maciel and Leite 2005; Shaw et al. 1980). Because of the many ways AOB can negatively impact normal life (including esthetics, self-esteem, speech, eating, etc.), many patients seek treatment for this malocclusion (Bailey et al. 2001).

Unfortunately, AOB is considered one of the most challenging malocclusions for orthodontists to treat and retain. Because the etiology of this malocclusion is complex and multifactorial, it is difficult to identify all contributing etiologic factors, and it is also hard to eliminate many of these skeletal, dental, respiratory, neurological, and habitual etiologic factors (Greenlee et al. 2011). Some of the main etiologic factors suggested to contribute to the development of AOB include digit or tongue habits, unfavorable growth patterns such as hyperdivergent skeletal patterns, and enlarged lymphatic tissue (Lentini-Oliveira et al. 2014). Both orthodontic and surgical treatment modalities are utilized in the treatment of AOB, and despite considerable experience with these strategies, there is no consensus on which methods are most successful. In 2015, the Adult Anterior Open Bite Study was launched by the National Dental Practice-Based Research Network (PBRN) to explore treatment recommendations, outcomes, and resultant vertical stability of adult AOB patients (Choi et al. 2018). One of the dentofacial characteristics examined in the study was the relative severity of the patients' AOB. A new index, the photographic open bite severity index (POSI), was developed to score severity of AOB based on an intraoral frontal photograph, an easily obtained photo that is routinely taken as part of orthodontic pre- and post-treatment records (Huang et al. 2019).

The POSI was developed to provide a quantitative assessment of AOB severity. While different AOB presentations are often described as *mild* or *severe*, these descriptors are subjective, non-standardized, and open to interpretation. An established index or score to quantitatively rate AOB severity is useful in many respects. Indices and scoring rubrics are commonly used in orthodontics; Little's Irregularity Index (LII), the Peer Assessment Rating (PAR) index, the Handicapping Labiolingual Deviation (HLD) index, the Index of Orthodontic Treatment Need (IOTN), and the ABO Discrepancy Index (DI) are examples of common indices used for different purposes in orthodontics (Brook and Shaw 1989; Cangialosi et al. 2004; Draker 1960; Little 1975; Parker 1998; Richmond et al. 1992). Establishing a standardized index or ranking system can help clinicians assess pre-treatment, post-treatment, and retention characteristics in an objective manner. Other indices can be used for epidemiologic studies, and some can even aid public health and insurance capacities to determine treatment priority. A valid and reliable score for AOB severity would be helpful as a standardized measure in clinical research, which might eventually be used as a common outcome for performing meta-analyses. It could also be helpful as a standardized measure for open bite in determining treatment need in the public health and insurance realms. Lastly, it could also be used to clarify diagnostic communication between clinicians.

Currently, open bite severity is often measured by the millimeters of negative overbite present either upon clinical examination or upon radiographic examination of a lateral cephalogram. But both measures only account for the vertical millimetric severity of open bite and do not account for the number of teeth affected, which also impacts severity. The clinical measure of open bite in millimeters can also be inconsistent due to differences in definition; some clinicians define it as the measurement from the incisal edge of the maxillary central incisors to the incisal edge of the lower anterior teeth with the posterior teeth in maximum contact, but others describe it as the vertical measurement between any opposing incisal edges of anterior teeth at which the greatest open bite presents (Fields et al. 2019; Parker 1998).

The radiographic measure of open bite on a lateral cephalogram can underestimate the vertical severity of an open bite because it only records the open bite at the most anterior point of the most anterior or prominent central incisors, due to the nature of the lateral cephalogram as a 2-dimensional representation of a 3-dimensional subject. Additionally, there is the concern of exposing patients to ionizing radiation when obtaining lateral cephalograms, which is appropriate when it is expected to provide information that can affect treatment and diagnosis, but inappropriate in other situations, such as when evaluating post-treatment retention with no intent to intervene (Abdelkarim and Jerrold 2018). Moreover, there have been attempts to categorize AOB as mild, moderate, or severe based on millimeters of vertical open bite either clinically or radiographically, but cut-off measurements for the three categories have varied in different studies and is not standardized (Agbaje et al. 2012; Tsang et al. 1997).

The POSI ranks AOB into 6 categories based on the specific teeth that have vertical overlap in an intraoral frontal photograph (Figure 1). A POSI score of 1 is given when there are 1 or 2 maxillary (Mx) lateral incisors without vertical overlap (but both Mx central incisors have vertical overlap); POSI score of 2 when there is 1 Mx central incisor without vertical overlap, but the other Mx central has vertical overlap; POSI score of 3 when there are 2 Mx central incisors without vertical overlap, but at least one Mx lateral has vertical overlap; POSI score of 4 when all 4 Mx incisors are without vertical overlap; POSI score of 5 when all anterior teeth are without overlap (canine to canine); and POSI score of 6 when POSI score 5 criteria are met plus at least 1 premolar without vertical overlap (Huang et al. 2019). The POSI score provides the missing component to assessing AOB severity by accounting for the number of teeth involved. Additionally, it is scored based on an intraoral frontal photograph, which is routinely obtained as part of pre- and post-treatment orthodontic records, and can be easily obtained at post-treatment retention checks without exposing patients to unnecessary radiation.

Figure 1. The 6 categories of the photographic open bite severity index (POSI), based on the type and number of teeth that do not have vertical overlap. Adapted from (Huang et al. 2019).

While many authors have described their own criteria for an ideal occlusal index, for any index to have merit for use, it must be established as valid and reliable (Carlos 1970; Liu et al. 2017; Shaw et al. 1991). Validity is the capacity of an index to measure what it claims to measure. In the case of the POSI, there are 2 tests of validity; the first is whether the POSI score correlates with a subjective clinical ranking of AOB severity, and the second is whether the POSI score correlates with a subjective clinical ranking of AOB treatment difficulty; both based solely on an intraoral frontal photograph. For an index to be reliable, it must be reproducible between different examiners (inter-examiner reliability), as well as by the same examiner at different timepoints (intra-examiner reliability). The POSI score must also demonstrate reliability (a consistent score) when the intraoral frontal photograph is taken at different vertical angles that deviate modestly from precisely parallel to the occlusal plane because such deviations routinely occur.

The purpose of this study was to assess the validity and reliability of the photographic open bite severity index (POSI).

AIMS

1) Evaluation of the reliability of the POSI score between examiners and within the same examiners.
2) Determination of the angular deviations within which the POSI score is reliable, based on the upward or downward tipping of the occlusal plane in the intraoral frontal photograph.
3) Evaluation of the validity of the POSI score by comparing correlations between POSI score and clinical evaluations of AOB severity and treatment difficulty.
4) Evaluation of whether a combination of the POSI score and a vertical millimetric measure of AOB correlates better with clinical evaluations, such as AOB severity or treatment difficulty than either measure alone.

METHODS

SAMPLE

The sample consisted of de-identified frontal intraoral photographs from open bite patients that are housed in the University of Washington Orthodontic Clinic database.

APPARATUS AND PROCEDURES

Examiner Evaluation Studies

To test the validity of the POSI score as an indicator of AOB severity and treatment difficulty, pre-treatment intraoral frontal photos from 84 anterior open bite patients were evaluated (14 in each POSI category). Case selection was completed as follows:

- The primary investigator viewed anonymized intraoral frontal photos and selected the first 14 photos of acceptable quality in each of the 6 POSI categories.
- Photo of acceptable quality: photo must have been captured from an angle ≤10 degrees from the occlusal plane (occlusal plane deviation angle). The occlusal plane deviation angle was approximated via digital measurement based on the averaged left and right angles formed by a horizontal line in the vertical midpoint of the open bite, and a second line that bisects the horizontal line from the maxillary midline to the mesiobuccal cusp of the maxillary first molar (Figure 2). Measurements were obtained by the primary investigator and a randomly selected subset of repeats were measured to test reliability.

Figure 2. Measurement of occlusal plane deviation in intraoral frontal photograph, measured from a horizontal line in the vertical midpoint of the open bite (pictured in yellow), and a second line that bisects the horizontal line from the maxillary midline to the mesiobuccal cusp of the maxillary first molar (pictured in white) on both the left and right sides.

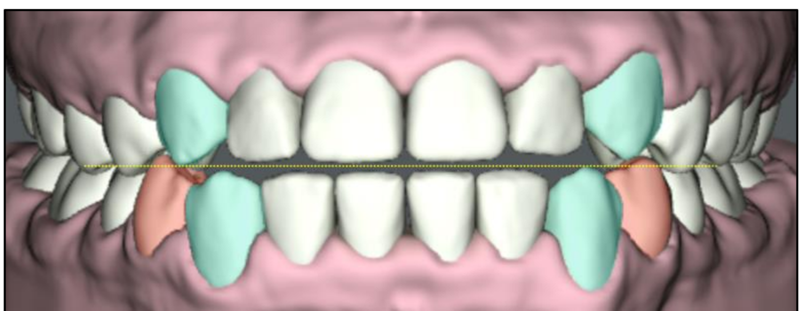
A panel of 5 practicing orthodontists was recruited from the University of Washington Department of Orthodontics faculty to serve as examiners. Inclusion criteria for taking part in the panel included 5 or more years of orthodontic practice experience. Exclusion criteria was previous involvement in research pertaining to the POSI. The panel consisted of 3 male examiners and 2 female examiners. There was a variety of educational backgrounds with training from 5 different dental schools and 3 different orthodontic programs.

Examiners completed three rating sessions individually using an online survey form: first they rated cases on severity of AOB, second, they rated cases on difficulty of AOB orthodontic treatment, and third they assigned cases a POSI score. The examiners were presented with only the intraoral frontal photographs and no other case information/records were provided. The order of the cases was randomized for all examiners at each session. For the first session, each examiner individually scored the cases based on their subjective impression of the clinical severity of AOB (Likert scale of 1-10 with 1 being the least severe AOB and 10 being the most severe AOB). For the second session, each examiner individually scored the cases based on their subjective evaluation of the clinical difficulty of AOB treatment (Likert scale of 1-10 with 1 being the lowest difficulty of treatment and 10 being the highest difficulty of treatment). For these first two sessions, examiners were asked to score only based on the AOB/vertical discrepancy and to not factor in other malocclusion elements they could discern from the intraoral frontal photographs such as anteroposterior factors, transverse discrepancies, restorative needs, periodontal concerns etc. For the third session, examiners were first trained and calibrated on POSI scoring in a brief (approximately 15-minute) session and provided with an instructional document detailing the POSI scoring categories, along with example photos of each category. Both the training and the instructional document outlined newly developed considerations regarding POSI scoring to improve standardization in scoring (Table 1). The panel subsequently completed the online survey form individually to assign each case a POSI score. Each session was completed a minimum of 2 weeks apart to minimize familiarity with the cases.

Table 1. POSI scoring considerations

Considerations for POSI scoring

1. Assume the patient is photographed in maximum intercuspation.
2. When looking for teeth not in "overlap," no portion of the incisal edge may overlap or appear to "touch" the opposing teeth.
3. For POSI 1-4, only consider the incisors. E.g., if a maxillary lateral incisor is overlapping with a mandibular canine, this does not count as "overlap."
4. For POSI 5/6 make vertical assessments of pairs of teeth, i.e., assess maxillary canine vertical overlap with the mandibular, the first premolar with first premolar, etc.
    a. Don't count "overlap" of a maxillary canine with a mandibular first premolar, don't count "overlap" of a maxillary first premolar with a mandibular second premolar, etc.
5. Visualize an imaginary plane to see if there is "overlap" when the teeth are not directly on top of one another in the photo.



POSI 5: assume canines oppose canines.
No "overlap" at canines using imaginary plane.

At each session, a total of 102 photos were scored in random order. This consisted of 84 different images (14 cases of each POSI category), and another 18 repeated photos (random selection of 3 of each POSI category) to assess intra-examiner reliability. The examiners' POSI, severity of AOB, and AOB treatment difficulty scores were assessed to determine correlations as well as to determine inter- and intra-examiner reliability. The "gold standard" POSI scores for each case were created by consensus among 4 experts who were involved with the development of the index (LT, SF, GH, GG) (Huang et al. 2019). The gold standard scores were compared to the raters' POSI scores to determine accuracy. The raters' AOB severity and treatment difficulty scores were compared to their POSI scores to assess POSI validity.

Angulation Reliability Studies

For the angulation reliability studies, 6 sets of Angle Class I study models were used: 1 of each POSI score (when oriented parallel to the occlusal plane), with varying degrees of millimetric open bites (Figure 3). Study models were digitally designed on SureSmile (Dentsply Sirona Inc.) software and then 3D printed in maximum intercuspation. Models were photographed from 30 degrees below the occlusal plane to 30 degrees above the occlusal plane in specified increments (±5, 10, 15, 20, and 30 degrees) at a distance of 0.5 meter away (typical focal distance for intraoral photos). Angular rotation was performed with a tripod and verified with a Klein Tools 935DAG digital angle gauge and level (Klein Tools, Inc.). The vertical midpoint of the lens was set at the same height as the midpoint of the open bite. A Canon EOS 200D DSLR camera equipped with a Canon EF 100mm f/2.8L Macro IS USM Lens and Canon Macro Ring Lite MR-14EX (Canon U.S.A., Inc.) was used to capture the photographs.

Figure 3. 3D-printed models of POSI 1-6.

Photographs were randomized into 4 groups. Each group of randomized photos was scored at separate time points with at least 2 weeks of separation time between each scoring session. Three raters were recruited for the angulation reliability studies: one orthodontic graduate student, one dental student, and one orthodontic faculty member involved in the development of the index. Using the photographs, POSI scores were rated independently by the two initial raters. Disagreements were discussed, and if a consensus could not be reached, the third examiner independently rated the case, and the majority POSI score was used.

Angulation Validity Threshold

Pretreatment intraoral frontal photos with the occlusal plane deviated >10-15 degrees were compared with those with less deviation to determine a threshold for the validity of the POSI score. This threshold would be used to identify intraoral frontal photographs that were more likely to result in erroneous POSI scores.

Combined Measurements (POSI and vertical measure of AOB)

The correlations between the combination of POSI score and vertical millimetric measures of open bite with AOB clinical evaluations (AOB severity and treatment difficulty ratings) were assessed to evaluate whether a combination would yield a stronger correlation with AOB clinical evaluations than POSI or millimetric measure of open bite would alone. Two versions of the combined POSI and vertical measure of AOB were studied - one using the millimetric measure of open bite as measured on a lateral cephalogram (ceph-measured OB), and one derived from a measurement based on the greatest vertical open bite as shown in the intraoral photograph (estimated clinical open bite), not restricted to the most anterior point of the central incisors, as would be the case with a lateral cephalogram. The latter version simulated vertically measuring the open bite at the most severe point, clinically, since that is clinical information (millimetric open bite at most severe point, as measured from clinical exam).

The ceph-measured OB was obtained by landmarking points in Dolphin, and then exporting the calculated negative overbite. The estimated clinical open bite (OB) was acquired by the primary investigator, who viewed the pretreatment intraoral frontal photos from the 84 cases included in the study, and visually estimated the greatest vertical opening in millimeters for each case. To test for intra-rater reliability, 18 randomly selected photos were repeated. Two tests were performed to test for validity of the estimated clinical OB. First, the method of estimating millimeters of opening from viewing the intraoral frontal photos was tested. To do so, a second estimation was obtained – the vertical opening in millimeters at the most prominent central incisors (where a lateral cephalogram measures OB) was estimated by viewing the intraoral frontal photo for the same 18 repeated photos, and then compared to the ceph-measured OB. Second, a ratio was created using digital measurements from the photos of the vertical opening at the most prominent central incisors (where a ceph measures OB) divided by the vertical opening between any opposing incisal edges of anterior teeth at which the greatest open bite presented (where estimated clinical OB was acquired). A ratio was used to bypass magnification issues present in measurements attained from photographs. The ceph-measured OB was divided by the ratio and the quotient was compared to the estimated clinical OB. Four photos were excluded due to photographed positive overbite or 0mm overbite at the most prominent central incisors, resulting in only 14 of the 18 repeated photos being used for both tests.

Combined Open Bite Severity Index (COSI)

Linear regression-based formulas were developed to quantify the combined measurements of POSI score and vertical millimetric measure of AOB for practical utilization. This combined open bite severity index (COSI) score would provide a single value that considers both the specific teeth affected (POSI) and the vertical degree of opening for use in similar applications that the POSI is applicable: as a standardized measure in research and in treatment need determination.

Data Analysis

A sample size of 84 cases and 5 examiners was chosen based on calculations for the inter-examiner reliability and the correlation of POSI score to AOB severity and treatment difficulty. Eighteen cases were repeated based on calculations for intra-examiner reliability testing. Based on ½ width of a 95% confidence interval for the kappa statistic of inter-examiner reliability, with at least 55 cases, the precision to estimate the kappa statistic for POSI, AOB severity, and AOB treatment difficulty was at least ±0.10 assuming a kappa statistic of 0.7 to 0.8. Based on ½ width of a 95% confidence interval for the kappa statistic of intra-examiner reliability, with 15 cases the precision to estimate the kappa statistic for POSI, AOB severity, and AOB treatment difficulty was ±0.10 assuming a kappa statistic of 0.9. Based on ½ width of a 95% confidence interval for the rank correlation, with 80 to 90 cases the precision to estimate the rank correlation between POSI and AOB severity and treatment difficulty was ±0.10 to 0.15 assuming a rank correlation of 0.7 to 0.8.

The kappa statistic and weighted kappa statistic along with 95% confidence intervals were used to evaluate the inter- and intra-examiner reliability of the POSI, AOB severity, and AOB treatment difficulty scores (Altaye et al. 2001). The weighted kappa allows for disagreements to be weighted differently, which is useful for ordered measures, such as the POSI, in order to weigh disagreements differently by how many values off from each other they are. A kappa or weighted kappa > 0.75 generally indicates excellent agreement, 0.40 to 0.75 fair to good agreement, and poor agreement below 0.40 (Fleiss 1981). Inter-examiner reliability was based on all 84 cases and intra-examiner reliability was based on a subset of 18 cases. The median and interquartile range were calculated by examiner for the POSI, AOB severity, and AOB treatment difficulty scores to assess for systematic differences among the examiners and between repeated ratings by the same examiner. Friedmann's test, a nonparametric test similar to repeated measures of ANOVA, was used to test for differences among the examiners and between repeated ratings by the same examiner (Kendall 1970).

The Spearman rank correlation was used to compare the POSI score to the AOB severity and to the AOB treatment difficulty scores to evaluate the validity of the POSI score. Analysis was completed using both the examiner-assigned POSI scores and the gold standard POSI scores. The rank correlation of AOB severity and POSI score was based on a linear regression of the ranked values for AOB severity on the ranked POSI scores and adjusted for the examiner. The bias-corrected and accelerated bootstrap method with 1000 replications was used to compute a 95% confidence interval for the rank correlation. The same approach was used to estimate rank correlations with estimated and measured overbite and for rank correlations with the AOB treatment difficulty. Correlations based on a single examiner were computed using a similar approach, except no adjustment for the examiner was necessary.

To assess the effect of angulation on the reliability of the POSI, 6 models (1 model with each of the possible POSI) were photographed from 30 degrees below the occlusal plane to 30 degrees above the occlusal plane in specified increments, ±5, 10, 15, 20, and 30 degrees, (11 different angles). The kappa statistic and weighted kappa statistic along with 95% confidence intervals were calculated for each angle to evaluate how the reliability varies by the degree of angulation and to establish a range of angulation that provides sufficient reliability. All analyses were performed using R statistical software (Team 2013).

All POSI scores used for combined measurements were gold standard POSI scores and not examiner-assigned POSI scores. The rank correlation of the combined measurements (POSI score plus either estimated clinical or ceph-measured open bite) was based on a linear regression of the ranked values for AOB severity on the ranked POSI scores and ranked overbite values, and adjusted for the examiner. The bootstrap method was used to compute a 95% confidence interval for the rank correlation, as well as to compute a 95% confidence interval for the difference between the rank correlations for POSI and combined measurements to evaluate whether the correlation was stronger for combined measurements than for POSI alone. The same approach was used for rank correlations with the AOB treatment difficulty. Correlations based on a single examiner were computed using a similar approach, except no adjustment for the examiner was necessary.

For the two tests used to evaluate the validity of the estimated clinical open bite (OB), the mean and standard deviation (SD), the mean (SD) of the differences and 95% confidence interval (CI) for the mean difference, the intraclass correlation coefficient (ICC) and 95% CI for the ICC, Dahlberg’s error, and the minimum and maximum for the absolute value of the difference between the two measurements were computed for each set of OB measurements compared.

Linear regression was used to estimate equations for different exploratory variations of the COSI using the AOB severity and treatment difficulty scores based on POSI only, POSI plus estimated clinical OB, and POSI plus ceph-measured OB. COSI equations were based on the gold standard POSI. The R-squared and root mean squared error (MSE) were used to describe the equation performance. R-squared describes how much of the variation in AOB severity and treatment difficulty can be explained by the equation. The root MSE is the square root of the average of square errors. E.g.,

$$\text{Root MSE} = \sqrt{\sum(severity - predicted\ severity)^2 / n}$$

If the prediction errors are normally distributed, then 68% of the calculated scores would be within ± 1 root MSE of the examiners' scores and 95% of the calculated scores would be within ± 2 root MSEs of the examiners' scores. The bias-corrected and accelerated bootstrap method and 5-fold cross-validation using the 5 examiners to partition the data were used to construct 95% confidence intervals for R-squared and root MSE.

RESULTS

EXAMINER EVALUATION STUDIES

Intraoral frontal photos from all 84 cases used were all determined to have been photographed without a significant deviation (± 10 degrees) from the occlusal plane. The intra-rater reliability for the occlusal plane deviation angle measurements was excellent, with an ICC of 0.99 (95% CI, 0.98 to 1.00) (Appendix 1).

AOB Severity Survey - Agreement & Reliability

The agreement for clinical evaluation of AOB severity for the cases was excellent amongst the 5 examiners on the panel (ICC = 0.76 (95% CI, 0.69 to 0.82) (Appendix 2&3). Each of the examiners also showed excellent intra-examiner reliability, with ICCs between 0.89 and 0.97 for all examiners (Appendix 4).

AOB Treatment Difficulty Survey - Agreement & Reliability

The agreement for clinical evaluation of AOB treatment difficulty for the cases was good amongst the 5 examiners on the panel (ICC = 0.67 (95% CI, 0.58 to 0.75) (Appendix 5&6). Each of the examiners also showed excellent intra-examiner reliability, with ICCs between 0.78 and 0.95 for all examiners (Appendix 7).

POSI Survey - Reliability & Accuracy

The POSI score reliability was excellent for all 5 examiners on the panel, with intra-examiner reliability ICCs between 0.93 and 0.96 for the 5 examiners (Table 2). The inter-examiner reliability ICC amongst the 5 examiners was excellent as well at 0.95 (95% CI, 0.94 to 0.97). All 5 examiners also showed excellent accuracy in POSI scoring, with weighted kappas between 0.90 and 0.96 when matched up with the gold standard POSI scores assigned to each case (Table 3).

Table 2. POSI Intra-examiner ICC, n=18

	Examiner				
	#1	#2	#3	#4	#5
Intra-examiner ICC	0.95	0.94	0.96	0.96	0.93
95% CI	0.63, 0.99	0.62, 0.98	0.75, 0.99	0.78, 0.99	0.60, 0.98
Weighted kappa	0.97	0.94	0.94	0.94	0.94
95% CI	0.91, 1.00	0.86, 1.00	0.86, 1.00	0.86, 1.00	0.86, 1.00
Unweighted kappa	0.93	0.87	0.87	0.87	0.87
95% CI	0.80, 1.00	0.69, 1.00	0.69, 1.00	0.69, 1.00	0.69, 1.00
Percent agreement (%)	94.4	88.9	88.9	88.9	88.9
95% CI (%)	72.7, 9.99	65.3, 98.6	65.3, 98.6	65.3, 98.6	65.3, 98.6

Table 3. POSI examiner scores as compared to the gold standard POSI, n=84

	Examiner				
	#1	#2	#3	#4	#5
Weighted kappa	0.90	0.96	0.91	0.94	0.95
95% CI	0.85, 0.95	0.92, 0.99	0.86, 0.96	0.91, 0.98	0.92, 0.98
Unweighted kappa	0.79	0.91	0.81	0.87	0.89
95% CI	0.69, 0.88	0.85, 0.98	0.72, 0.91	0.79, 0.95	0.81, 0.96
Percent agreement (%)	82.1	92.9	84.5	89.3	90.5
95% CI (%)	72.3, 89.6	85.1, 97.3	75.0, 91.5	80.6, 95.0	82.1, 95.8

POSI Discrepancies

Eleven of the 84 total cases were identified as having the most frequent discrepancies between examiner-assigned POSI score and gold standard POSI score: 4 cases showed 3 counts of discrepancies, and 7 cases showed 2 counts of discrepancies (Table 4).

Table 4. Discrepancies between examiner-assigned POSI scores and gold standard POSI scores

Discrepant Case #	# of Discrepancies	Gold Standard POSI	Discrepant Examiner POSI scores
1	3	5	4, 4, 4
2	3	4	3, 5, 6
3	3	3	4, 4, 4
4	3	6	5, 5, 5
5	2	5	4, 4
6	2	5	4, 4
7	2	6	5, 5
8	2	5	4, 4
9	2	2	1, 1
10	2	6	5, 5
11	2	3	4, 4

POSI Validity

Correlations were studied using both the gold standard POSI scores and the examiner-assigned POSI scores. The correlations between gold standard POSI scores and both AOB severity and treatment difficulty scores were moderate, with Spearman rank correlation coefficients of 0.71 and 0.66, respectively (Appendix 8). Scatter plots illustrate these relationships in Figure 4. The correlations between examiner-assigned POSI scores for all examiners combined and both AOB severity and treatment difficulty scores were similar to the correlations found with gold standard POSI scores (Appendix 9), with Spearman rank correlation coefficients of 0.73 and 0.66, respectively (Appendix 10).

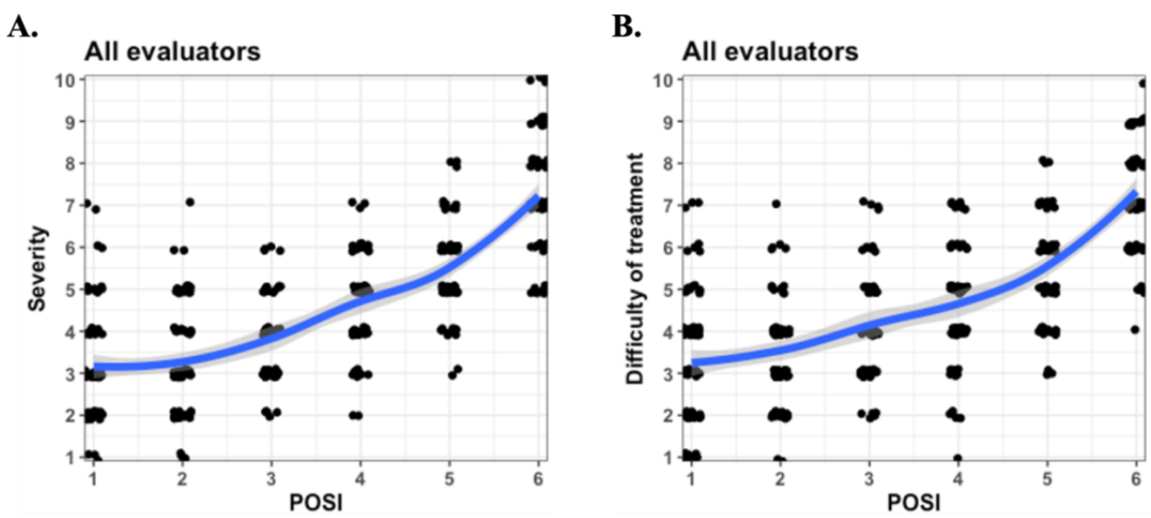


Figure 4. Scatter plots of A) AOB Severity versus gold standard POSI and B) AOB Treatment Difficulty versus gold standard POSI

There was also a moderate correlation between the AOB severity and treatment difficulty scores, with a Spearman rank correlation coefficient of 0.66 (95% CI, 0.61 to 0.71) (Appendix 11). Figure 5 shows a scatter plot of AOB severity vs. AOB treatment difficulty scores, depicting the positive correlation.

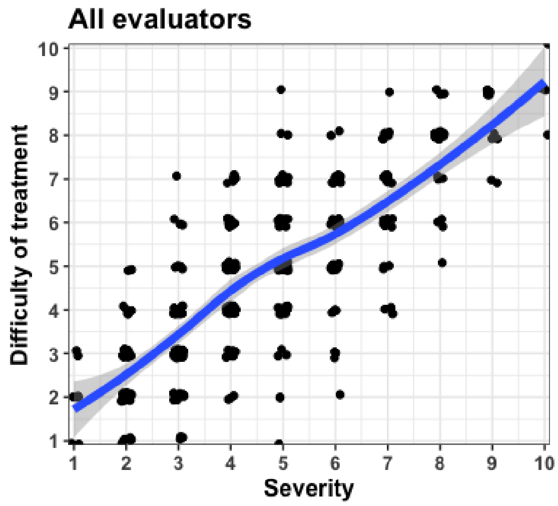


Figure 5. Scatter plot of AOB Treatment Difficulty vs. AOB Severity

ANGULATION RELIABILITY STUDIES

Six models with a true POSI score of 1, 2, 3, 4, 5, and 6 (Figure 3) were each photographed at angles of -30, -20, -15, -10, -5, 0, 5, 10, 15, 20 and 30 degrees. A model's true POSI score indicates its photographed POSI score with 0° of deviation from the occlusal plane. Negative angles denoted a photo captured from below, and positive angles denoted a photo captured from above (example shown in Figure 6). Photos were scored by the two raters (an orthodontic graduate student and a dental student), and the two raters assigned the same POSI score to all photos except for 4 instances. The 4 photos with disagreement were a true POSI of 6 at angles of +15, +20, and +30 degrees (rater 1 giving a score of 6 and rater 2 giving a score of 5), and a true POSI of 4 at an angle of -15 degrees (rater 1 giving a score of 4 and rater 2 giving a score of 3).

In cases of disagreement, the third rater gave a POSI score, which agreed with the orthodontic graduate student, and not the dental student, in all four cases. The consensus POSI score was the value of the two identical scores for each scenario. Consensus POSI scores and percent agreement are reported in Table 5. The absence of an AOB was recorded as a POSI 0.

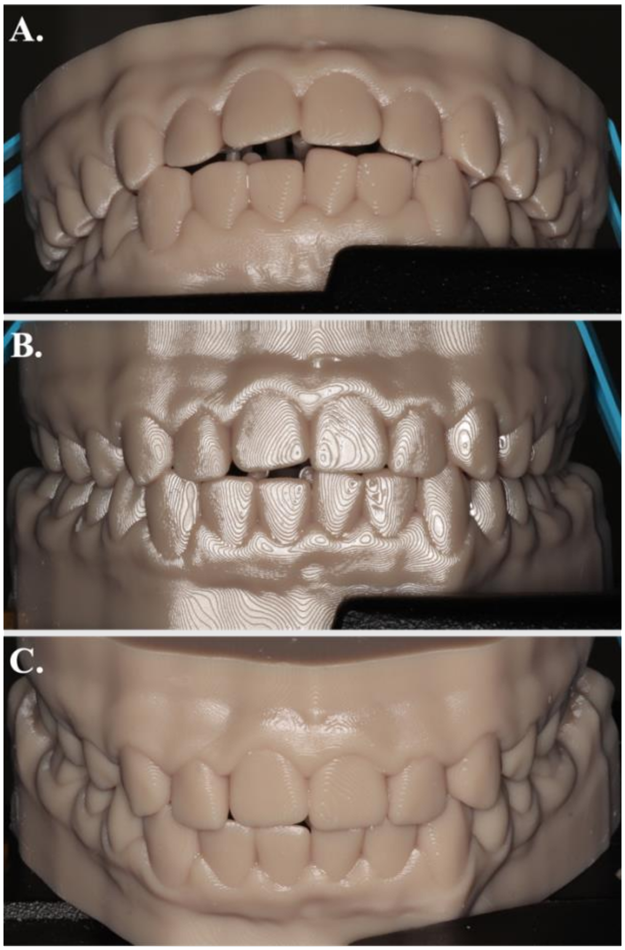


Figure 6. Example of positive vs. negative angle: true POSI 2 model photographed at A) -20 degrees, B) 0 degrees, C) +20 degrees. At -20 degrees, the true POSI 2 model appears open at all incisors (photographs as POSI 4), and at +20 degrees the true POSI 2 model appears to have no open bite (photographs as POSI 0; no AOB).

Table 5. Consensus POSI scores by angle and percent agreement with true POSI; green denotes agreement with true POSI and red denotes disagreement with true POSI. Absence of AOB = POSI 0.

		True POSI Score						
		1	2	3	4	5	6	
	Angle (°)	Consensus Photographed POSI Score						% Agreement with True POSI
Photographed from below	-30	4	4	4	4	5	6	50.0%
	-20	1	4	4	4	5	6	66.7%
	-15	1	4	3	4	5	6	83.3%
	-10	1	2	3	4	5	6	100%
	-5	1	2	3	4	5	6	100%
	0	1	2	3	4	5	6	100%
Photographed from above	+5	1	2	3	4	5	6	100%
	+10	1	2	3	4	5	6	100%
	+15	1	0	3	1	5	6	66.7%
	+20	0	0	3	1	5	6	50.0%
	+30	0	0	2	0	5	6	33.3%

The results showed that the consensus POSI scores matched the true POSI scores for angles -10 to 10 degrees, and for true POSI scores of 5 and 6 at all angles. For angles of 15 to 30 degrees, true POSI scores <5 tended to be scored lower than the true score. For angles of -30 to -15 degrees, true POSI scores <5 tended to be scored higher than the true score.

Angulation Validity Threshold

Through studying the pretreatment intraoral photos, it was found that typically, if the angular deviation of the photograph was greater than ~10-15 degrees above the occlusal plane, then an imagined horizontal line in the vertical midpoint of the open bite, pictured in yellow in Figure 2, would oftentimes be apical to the clinical crowns of the mandibular first molars, and vice versa for ~10-15 degrees below the occlusal plane with the maxillary first molars. Therefore, when a horizontal line is imagined or drawn bisecting the anterior open bite, and it does not fall above or below the clinical crown of the first molars, assigning a POSI score should be valid. If the horizontal line does fall above or below the respective first molar clinical crown, another photograph should be obtained with a more favorable angulation for assessing the POSI score.

Combined Measurements

For the estimated clinical open bite (OB) - vertical measurement of open bite at the most severe tooth (not restricted to the most anterior point of the central incisors) estimated from intraoral frontal photos - the intra-rater reliability based on 18 randomly selected repeats was excellent, with an ICC of 0.98 (95% CI 0.96 to 0.99) (Appendix 12). The first test of validity for estimated clinical OB revealed an excellent ICC of 0.78 (95% CI, 0.46 to 0.92) when comparing the ceph-measured OB to the second estimate of OB from the intraoral photos – this time at the location where the ceph measures OB from (most prominent/central incisors) (Appendix 13). The second test of validity used the ratio created by using digital measurements from the photo of the vertical opening at the most prominent/central incisors (where a ceph measures OB) divided by the greatest vertical opening of all anterior teeth (where estimated clinical OB was acquired) (Appendix 14). The ceph-measured OB divided by the ratio was compared to the estimated clinical OB, resulting in a good ICC of 0.72 (95% CI, 0.33 to 0.90).

Combined Open Bite Severity Index (COSI)

Four potential COSI formulas were explored: 2 based on AOB severity scores and 2 based on AOB treatment difficulty scores, each with one variation using the estimated clinical OB and another variation using the ceph-measured OB (Table 6). Like the correlation patterns, the COSI variation using the POSI + estimated clinical OB with AOB Severity scores yielded the best result, with an R-squared value of 64.1%. This formula was the one designated as the definitive COSI formula:

$$COSI = 1.55 + 0.50*POSI + 0.39*OB$$

The R-squared values for the POSI + ceph-measured OB were the second highest, followed by the POSI only and estimated clinical OB only with similar R-squared values. The R-squared values for ceph-measured OB only were consistently the lowest.

Table 6. Development of the COSI:l inear regression-based formulas for A) AOB Severity and B) AOB Treatment Difficulty based on POSI only and COSI variations. In the "Formula" column, OB is the millimeters of open bite, aka negative overbite. The formulas should not be used in the case of positive overbite.

A.

	AOB Severity based on:	Formula	R-squared (95% CI)	Root MSE (95% CI)
	POSI only	1.83 + 0.79*POSI	51.4% (43.3 to 58.3)	1.31 (1.22 to 1.43)
	Estimated clinical OB only	2.53 – 0.63*OB	50.2% (42.2 to 57.1)	1.76 (1.53 to 2.06)
	Ceph-measured OB only	3.25 – 0.64*OB	42.5% (34.1 to 50.2)	2.04 (1.81 to 2.33)
COSI variations	POSI + Estimated clinical OB	1.55 + 0.50*POSI + 0.39*OB	64.1% (56.1 to 70.2)	1.13 (1.05 to 1.24)
COSI variations	POSI + Ceph-measured OB	1.95 + 0.56*POSI + 0.33*OB	58.2% (49.1 to 65.2)	1.22 (1.13 to 1.33)

B.

	AOB Tx Difficulty based on:	Formula	R-squared (95% CI)	Root MSE (95% CI)
	POSI only	2.03 + 0.77*POSI	46.5% (37.9 to 54.0)	1.41 (1.32 to 1.52)
	Estimated clinical OB only	2.67 – 0.62*OB	47.1% (40.3 to 54.1)	1.98 (1.75 to 2.28)
	Ceph-measured OB only	3.46 – 0.61*OB	36.1% (28.5 to 43.8)	2.39 (2.11 to 2.72)
COSI variations	POSI + Estimated clinical OB	1.74 + 0.48*POSI + 0.40*OB	59.2% (51.3 to 65.6)	1.23 (1.15 to 1.34)
COSI variations	POSI + Ceph-measured OB	2.14 + 0.56*POSI + 0.29*OB	51.6% (42.2 to 59.1)	1.34 (1.25 to 1.45)

A scatter plot of the AOB severity versus COSI is illustrated in Figure 7, demonstrating how cases with higher AOB severity ratings tend to score higher (closer to a COSI score of 10), and cases with lower AOB severity ratings tend to score lower (closer to a COSI score of 2).

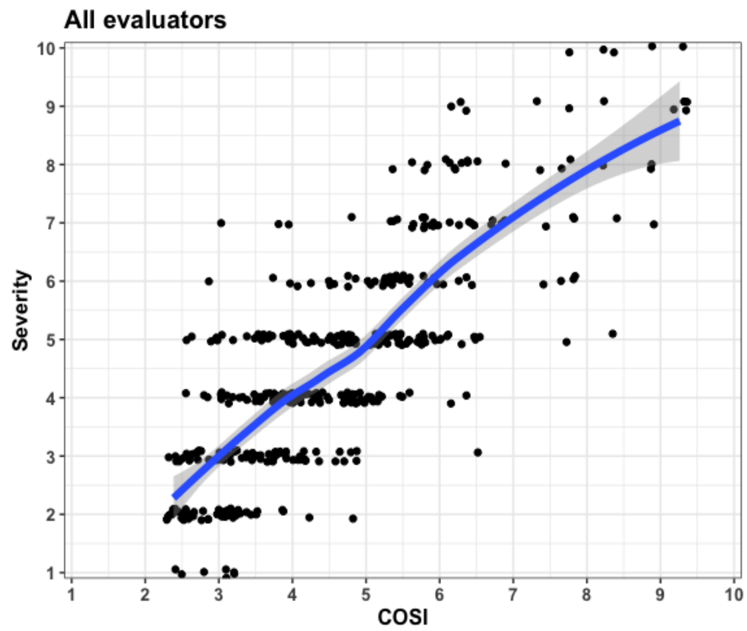


Figure 7. Scatter plot of AOB Severity vs. COSI. The COSI formula used was created based off gold standard POSI + estimated clinical OB.

DISCUSSION

Previous studies show that gathering ratings and judgments from an experienced panel of orthodontist examiners is an accepted standard in developing and validating orthodontic indices (Brook and Shaw 1989; Liu et al. 2017; Shaw et al. 1991). The examiners in this study showed excellent intra-rater reliability for survey questions, and high inter-rater reliability amongst the panel as well, suggesting that these orthodontists' clinical judgements were sound references for POSI validation and that the POSI is reliable. In fact, the intra-rater reliability for the POSI scoring was the highest of the three sets of scoring that the panel completed. This can be explained by the fact that the POSI score is obtained by a set of relatively objective rules, whereas one's clinical evaluation of AOB severity or treatment difficulty is more subjective and less structured. The panel's POSI scores were also highly accurate following only a brief (~15m) training session, making the weighted kappas ranging from 0.90 to 0.96 (versus gold standard POSI scores) even more demonstrative of how accessible it is to learn and utilize the POSI.

POSI DISCREPANCIES

Although the panel's POSI scoring was highly accurate overall, it is useful to scrutinize the cases with most frequent discrepancies between examiner POSI score and gold standard POSI score to understand the POSI's limitations. Looking at the 11 cases that the panel most often scored incorrectly, the most frequent mistake was scoring a gold standard POSI 5 as a POSI 4, which occurred in 4 cases (discrepant case #s 1, 5, 6, 8). In all 4 cases, the canines had no "overlap" when utilizing the imaginary plane, as described by the 5th point in Table 1, but this shows that it can be challenging to judge overlap based on an imaginary plane when the teeth are not directly on top of one another, which makes the differentiation between POSI 4 and POSI 5 difficult.

In 3 cases, a gold standard POSI 6 was scored as a POSI 5. The challenge of imagining a horizontal plane could also be at play here, but the more likely explanation for this mistake is that the examiner judged overlap between the maxillary first premolar against the mandibular second premolar, instead of correctly judging against the mandibular first premolar. As seen in discrepant case #4, pictured in Figure 8, the patient’s left first premolars are clearly not in overlap (along with the canine-canine segment) so this fulfills the criteria for a POSI 6, but if one were to accidentally judge the patient’s maxillary left first premolar against the mandibular left second premolar that appears directly below, in a frontal photo, then one could mistakenly label the case a POSI 5. This is an easy mistake to make, as the maxillary first premolar is often right on top of the mandibular second premolar in an intraoral frontal photo where AP differences are hard to discern. The first and second premolars do look alike, but this mistake can be prevented by more careful consideration whilst scoring.

Figure 8. Discrepant case #4: gold standard POSI 6 but was mistaken 3 times for a POSI 5.

Discrepant case #11 was likely also a case of scoring based off the tooth directly below the maxillary tooth instead of carefully considering the POSI rules and utilization of the imaginary horizontal plane (Figure 9.a). However, this presents a more interesting scenario because the mistake likely only developed due to a significant asymmetry between the maxillary and mandibular dentition. This was a case where a gold standard POSI 3 was rated as a POSI 4 twice because at first glance, it might have appeared to the examiners that the incisors were all in open bite. However, if one remembers to judge all the maxillary incisors against the mandibular incisors and also to use the imaginary horizontal plane for when the teeth to be evaluated are not directly above and below one another, it becomes clear that the right laterals are in "overlap" despite what looks like a mandibular deviation causing the mandibular left lateral to be significantly farther left from where one would typically expect to see it positioned in an ideal occlusal relationship (Figure 9.b). This illustrates the need for POSI raters to be diligent of deviations from expected occlusal relations and how these deviations can affect where teeth are positioned.

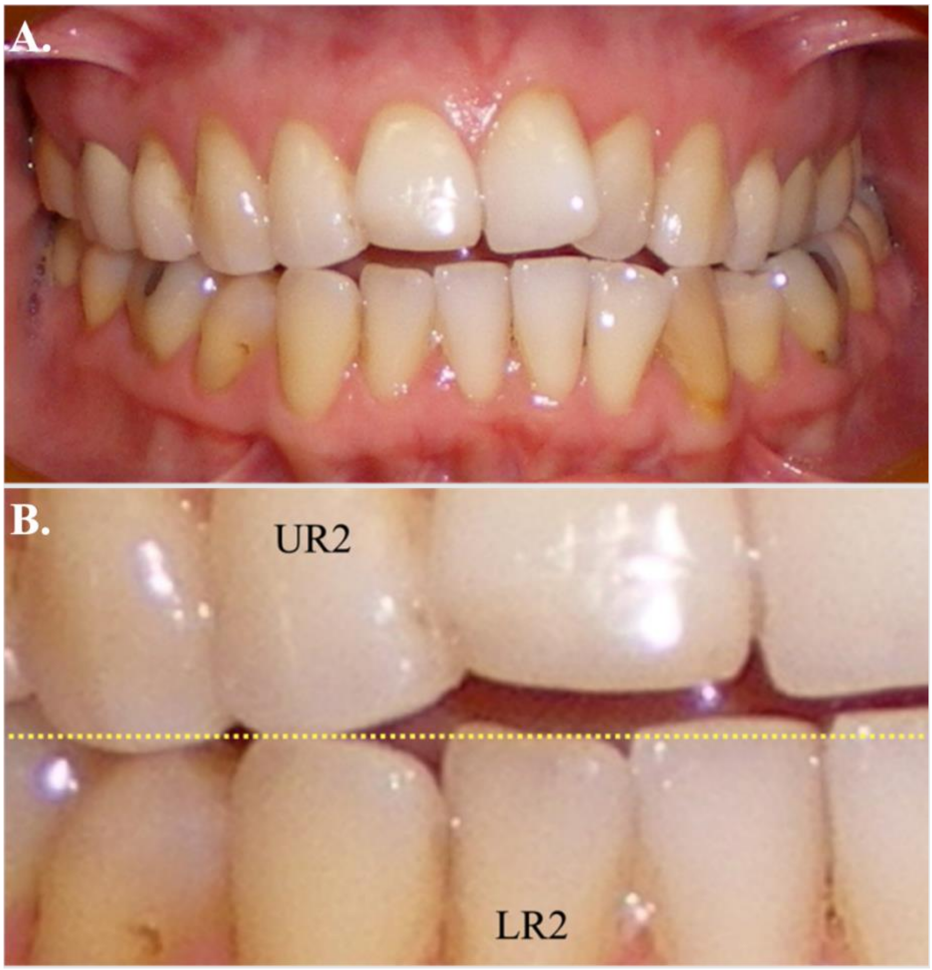


Figure 9. A) Discrepant case #11 with B) zoomed-in view of "overlap" between right lateral incisors utilizing imaginary horizontal plane (dotted yellow line).

Mistakes in POSI scoring were made in 2 cases (discrepant case #s 3 and 9) where it was difficult to determine whether two opposing teeth appeared to “touch” or not because they were so close together but not obviously overlapping. In discrepant case #3 (Figure 10), the gold standard score was POSI 3, but it was mistaken 3 times as a POSI 4 because of a set of right laterals that were barely appearing to “touch” and thus fulfill the “overlap” criteria. This situation represents another challenge in POSI scoring, as it is not always easy to determine if the teeth just barely “touch” and thus, meet the “overlap” criteria for POSI scoring as described in point #2 of Table 1’s POSI scoring considerations, or if the teeth do not “touch” and there’s a sliver of space pictured between the opposing teeth.

Figure 10. A) Discrepant case #3 with B) zoomed-in view of right lateral incisors that appear to "touch."

The last POSI scoring discrepancy was discrepant case #2 – a gold standard POSI 4 that was mistakenly rated 3 times: once as a POSI 3, once as a POSI 5, and once as a POSI 6 (Figure

11). If one were to accidentally overlook the “overlap” at the left canines, then it is conceivable how one would then rate this a POSI 6, but it is hard to say how one could mistakenly label this case a POSI 3 or POSI 5. It is possible to mistakenly judge the left central incisors as “touching” since the sliver of space between the two teeth is so slight, but that would result in a score of POSI 2, not POSI 3 or 5. It seems that possible explanations for the mistaken ratings of POSI 3 and POSI 5 could be that the examiners got the POSI categories mixed up in this single instance (despite having reference materials available), or that the examiners accidentally clicked the wrong POSI score (since POSI 3 and POSI 5 are adjacent to the gold standard score of POSI 4). If so, in a survey of 102 photos/examiner to assign POSI scores to and 5 total examiners, the occurrence of only 2 complete mistakes is a rather low number and a good overall outcome. However, it is surprising that both mistakes happened to occur on the same case.

Figure 11. Discrepant case #2: gold standard POSI 4 as all incisors are open but there is “overlap” of the left canines.

POSI Validity

For the POSI to be valid as a measure of AOB severity as the name claims, it must correlate with the panel's subjective clinical evaluation of AOB severity and treatment difficulty. This study found robust correlations between gold standard POSI scores and the panel's evaluations of AOB severity and treatment difficulty, and nearly identical correlations between the examiner-assigned POSI scores and their evaluations of AOB severity and treatment difficulty. This confirms that the POSI is a valid index, meaning it is a good reflection of AOB severity. The finding that the correlation coefficients were nearly identical when using gold standard POSI scores versus examiner-assigned POSI scores was to be expected since the examiner-assigned POSI scores so accurately reflected the gold standard scores.

AOB severity and AOB treatment difficulty are not two completely independent variables, as it makes sense that as severity increases, treatment difficulty does as well, and this is reflected in the marked correlation found between the panel's AOB severity and AOB treatment difficulty rank scores. However, while they are related, they are not synonymous, and the inter-examiner reliability ICC for AOB severity was higher than the inter-examiner ICC for AOB treatment difficulty. One possible explanation for why clinicians' impressions of AOB severity may be more standardized across the board than that of AOB treatment difficulty, is that treatment difficulty also depends largely on an individual's ideal treatment method. For example, a case of a certain severity would have very different levels of treatment difficulty depending on whether it is treated with orthognathic surgery or not. It is possible that an orthodontist who routinely treats surgical cases could ideate a surgical treatment plan and assign a lower AOB treatment difficulty score than an orthodontist who doesn't routinely treat surgical cases and may instead assign the same case a higher AOB treatment difficulty score while envisioning a nonsurgical treatment modality.

Variations in AOB Severity & Treatment Difficulty

When evaluating the AOB severity and treatment difficulty ratings by individual POSI categories in Figure 4, a sizeable spread of ratings was observed for each POSI score. This begs the question – what makes an AOB low vs. high severity/difficulty at the same POSI score? After assessing the photos with the lowest and highest AOB severity and treatment difficulty scores of each POSI category, a few trends were observed. The most consistent pattern seen was the influence of the vertical degree of open bite, with less opening in cases with lower AOB severity/treatment difficulty scores and more opening in cases with higher AOB severity/treatment difficulty scores when compared to others in the same POSI category (Figure 12). A secondary trend noted was the presence of less crowding in some cases with lower AOB severity/treatment difficulty scores and more crowding in some cases with higher AOB severity/treatment difficulty scores when compared to others in the same POSI category (Figure 12). Additionally, for the POSI 1 and 2 categories (categories requiring overlap at one or two maxillary central incisors), cases with the highest AOB severity/treatment difficulty scores in each POSI category had lack of overlap at both maxillary laterals and sometimes beyond (Figures 12.g&h). Lastly, anterior and posterior crossbites were seen in some cases with the lowest as well as the highest AOB severity/treatment difficulty scores for their POSI scores (Figures 12.b-e, g, i-l). Reflecting on these trends, it is logical that the presence of a smaller vertical open bite would reflect in a lower AOB severity or treatment difficulty score, and a larger vertical open bite or more teeth involved in the open bite would reflect in a higher AOB severity or treatment difficulty score. However, since examiners were instructed not to factor in elements of the malocclusion other than AOB/vertical discrepancy, it is interesting that more crowding was found in many cases with the highest AOB severity and treatment difficulty for their POSI category. One possible explanation is that it was difficult for examiners to completely ignore the crowding, and its presence could have affected the subjective clinical evaluations despite the examiners' seemingly capable ability to disregard the presence or absence of a crossbite. Another possibility is that the examiners considered crowding to be a component of the vertical discrepancy under the idea of a nonextraction treatment plan where resolving significant crowding would have a tendency to worsen open bites.

Figure 12. A-F) Photos among the lowest AOB severity & treatment difficulty ratings in order from POSI 1-6. G-L) Photos among the highest AOB severity & treatment difficulty ratings in order from POSI 1-6.

While the POSI has been found to have excellent inter- and intra-rater reliability in this study, one unique consideration that a photographic index of this type has is how reliable it is when the photo is taken from above or below the occlusal plane. It would be unreasonable to expect for clinicians to have the precision to consistently take photos exactly parallel to the occlusal plane, and minimal deviations are acceptable. However, it is important to define the limit at which a deviation becomes unacceptable due to resultant inaccuracies in the POSI score, e.g., an AOB photographed from far above the occlusal plane could appear closed, and a bite with overlap photographed from far below the occlusal plane has the potential to appear like an AOB is present.

In this study, it was found that when intraoral frontal photographs were captured with up to 10 degrees of vertical deviation (either above or below the occlusal plane), the photo would still reflect the same POSI score as the true POSI (POSI at 0 degrees deviation) (Table 5). This gives clinicians a 20-degree range in which minor angular deviations would still yield the true POSI. The photos obtained at +/- 15, 20, and 30 degrees all resulted in some discrepancies between the photographed POSI score and the true POSI score (Table 5).

For angles that were photographed from above (+15, +20, and +30 degrees), it was found that for true POSI scores <5, the photographed POSI score was lower than the true score if discrepant. This is expected, as when you photograph an AOB from high above, it will tend to look less open. The models designed for true POSI 5 and 6 photographed as their true POSI scores even at +15, +20, and +30 degrees, but they also had the greatest degree of vertical open bite of all the models, with open bites of 4.4mm and 3.6mm, respectively (Appendix 15). It is possible that had the opening been smaller in the vertical dimension for POSI 5 and POSI 6 models, then at +15, +20, and +30 degrees, the photographed POSI scores could also score lower than their true POSI scores. The true POSI 3 model also showed an interesting pattern of maintaining a POSI score of 3 even when photographed from +15 and +20 degrees, and only scoring as a lower POSI score when heavily angulated at +30 degrees. It is likely that this is because the true POSI 3 model was also designed with a large vertical open bite – the 3rd largest (only smaller than the POSI 5 and POSI 6 models) – which supports the explanation that the severity of vertical opening also affects the resultant POSI score when photographed from above the occlusal plane.

For angles that were photographed from below (-15, -20, and -30 degrees), it was found that for true POSI scores <5, the photographed POSI score was higher than the true score if discrepant. This is also expected, as when you take photos from below the occlusal plane, it can look like teeth are in open bite even when they are not, when viewed parallel to the occlusal plane, as long as there is sufficient overjet present. It is also prudent to keep in mind that the magnitude of overjet affects the photographed POSI when angled from below the occlusal plane because for a pair of teeth with positive overbite, the larger the overjet that exists between the pair, the smaller the negative angulation it takes for the pair to lose their overlap. In summary, as long as the vertical deviation of an intraoral photograph is limited to 10 degrees or less, above or below the occlusal plane, the POSI is most likely reliable (photographed POSI score = true POSI score). At 15 degrees above or below the occlusal plane, POSI scores were found to be reliable the majority of the time, but there are risks of discrepancies. At deviations of 20 or 30 degrees above or below the occlusal plane, the photographed POSI is discrepant about half the time.

In practice, it may not be practical for a clinician to digitally or manually, with a protractor, measure the angular deviation of their intraoral frontal photograph before deciding to retake it, and an easy way to identify whether the angulation is too deviated would be useful. Because it was found that an imagined horizontal line in the vertical midpoint of the open bite (like the yellow line pictured in Figure 2) would intersect the clinical crown margins of the mandibular first molars for a deviation of ~10-15 degrees above the occlusal plane, and vice versa for ~10-15 degrees below the occlusal plane with the gingival margins of the maxillary first molars, this line can be used as a rough rule of thumb to determine the limit for a well-angulated photo. This rule of thumb would apply best to adult patients (since it was developed based on photos of adult dentition that isn't undergoing substantial active eruption) without significant aberrations to the first molars (such as supraeruption, ankylosis, or significant A-P discrepancies). This guide could be applied if ever one were concerned regarding the angulation of a photo immediately after capture so a retake could be obtained expediently, if needed.

Combined Measurements and COSI

While the POSI categorizes anterior open bites based on the specific teeth affected, it does not account for the vertical measure of AOB - the magnitude of negative overbite. To be comprehensive, the correlation between POSI, vertical measure of AOB, and the combination of POSI and vertical measure of AOB with AOB severity and treatment difficulty scores were computed and compared. Since open bite is often measured as the millimeters of negative overbite either on clinical exam or on a lateral cephalogram, both measures were evaluated in this study. The estimated clinical open bite (OB) served as the measure for the millimeters of negative overbite for our purposes, and our analysis revealed sufficient reliability and validity for this approximation. All measures of POSI used in our correlation studies with combined measurements (POSI + vertical measure of AOB) were the gold standard POSI scores and not the examiner-assigned POSI scores. While the scores were nearly identical across the board, the gold standard POSI scores were used because, as described previously, there were some discrepancies between the two sets of POSI scores that were largely due to errors from the examiners in following the POSI scoring rules (Table 1), such as scoring against the incorrect opposing teeth, and not as much due to a subjective difference of clinical opinion.

To create a practical and easily utilized measure of AOB severity that considers both the POSI and the vertical degree of open bite, linear regression-based formulas were explored to create the combined open bite severity index (COSI). Of the 4 exploratory variations of the COSI detailed in Table 6, the formula developed using AOB severity scores based on POSI and estimated clinical OB (1.55 + 0.50*POSI + 0.39*OB) was the one formula recommended for use as the definitive COSI. It makes sense for the definitive COSI formula to utilize AOB severity scores instead of AOB treatment difficulty scores since the AOB severity scores were more standardized between examiners, and not as likely to be affected by variations in a practitioner's preferred treatment method. The COSI formula also combines both POSI score and estimated clinical OB (vertical measure of OB at location of greatest opening) to produce a superior R-squared value, which matches the finding that accounting for both these measures yields the best correlation with subjective clinical evaluations of AOB severity. The definitive COSI formula also has the highest R-squared value compared to the other exploratory COSI variations, and this value indicates that 64.1% of the variation in AOB severity can likely be explained by this combination of the POSI score and vertical measure at the location of greatest open bite.

The combination of POSI + estimated clinical OB was able to explain the largest percentage of variation in AOB severity. The ceph-measured OB, however, explained the lowest percentage of variation in AOB severity. The correlations with AOB treatment difficulty followed the same pattern as with AOB severity. This elucidates that while the POSI is a valid measure of AOB severity and treatment difficulty, accounting for both the POSI and vertical degree of open bite (with the estimated clinical OB) does improve the approximation of AOB severity and difficulty of treatment. Based on these results, it is recommended that either the POSI, the largest vertical open bite of the anterior teeth (in mm), or the combination of the 2 (the COSI), be used as valid measures to characterize anterior open bite for future research or clinical evaluation of treatment severity or need. Of note, it also appears that the ceph-measured OB alone provides the lowest correlation with AOB severity and treatment difficulty of all the measurements investigated in this study, yet it is one of the most common metrics used in AOB research studies. This low correlation is likely due to the limitation of the radiographic measure of open bite on a lateral cephalogram to only measure the open bite at the most anterior opening (typically at the central incisors). As a result, if the location of the greatest vertical opening is anywhere other than the exact spot that is most anterior on a lateral cephalogram (e.g., at the lateral incisors) then the measurement would not accurately represent the severity of the vertical opening. Based on this finding, it would be preferred to use the COSI, POSI, or millimetric open bite measured at the greatest vertical opening over the ceph-measured OB alone in future AOB studies or evaluations of treatment need.

LIMITATIONS

Our panel of 5 is relatively small but increasing the number would have been challenging given the commitment, scheduling, and experience requirements necessary. Because all panel members were faculty members at a single institution (University of Washington Department of Orthodontics), there may exist similarities in orthodontic diagnosis and treatment styles so there could be concern that their opinions are not generalizable to orthodontists outside of the University of Washington. However, the panel comes from a diverse assortment of dental and orthodontic training backgrounds. Similarly, all panel members practice orthodontics in the greater Seattle, WA area, and since there can exist geographic influences on orthodontic assessments (Richmond and Daniels 1998), there could be concern that the panel is not representative of orthodontists elsewhere. However, the panel did consist of examiners with clinical experience domestically from Washington, California, and Michigan, and internationally from Canada and Iran.

There did exist discrepancies in POSI scoring between the panel of examiners and the gold standard POSI scores. However, there were very few discrepancies and the accuracy was very high at 88% (range: 82-93%). This was after a single ~15-minute training/calibration session, so if the number of calibration sessions were to be increased, it would be expected that the accuracy would be even greater. Out of the sparse discrepancies, most errors seemed to relate to referencing the correct teeth in judging the presence or absence of overlap. While it may sometimes be difficult for an evaluator to remember to judge incisors against incisors in POSI 1-4, to only judge canines and posterior teeth against their matching opposite-arch counterpart for POSI 5 and 6, and to visualize an imaginary plane when the teeth are not photographed directly on top of one another, this is a limitation that we accept because we want to be reliable: e.g., a maxillary first premolar may not actually be opposing a mandibular first premolar depending on the anteroposterior relationship, but deciphering which teeth actually oppose each other from an intraoral frontal photo is challenging so it is more reliable to instead abide by these rules.

Regarding the examiners' AOB severity & treatment difficulty ratings, it is difficult to say whether they were truly only accounting for the anterior open bite/vertical discrepancy, as instructed. While the instructions clearly stated to ignore other elements of the malocclusion, it is possible that one could have missed that detail, could have forgotten to do so, or could have subconsciously been affected by other elements of the malocclusion visible in the photo.

The POSI uses the specific teeth involved in an AOB to index severity but does not comprehensively factor in every element that affects AOB severity and treatment difficulty. That is why the COSI provides a better picture with AOB severity and treatment difficulty. However, the COSI only accounts for about two thirds of the variation in AOB severity in our study, so there are still other factors that play a role in practitioners' subjective evaluations of AOB severity. Based on the trends seen in this study, crowding could possibly account for some of that remaining variability, but future studies are needed for confirmation. There are additional important measures, other than crowding, that were also not investigated in this study, such as mandibular plane angle. Many orthodontists would agree that a steep mandibular plane angle, as measured on a lateral cephalogram, is a factor that increases the severity and treatment difficulty of an anterior open bite. However, to encapsulate every element that relates to AOB severity into an index would be arduous and would also require a much more cumbersome rating system. Plus, both the POSI and COSI already account for an important aspect that current measurements do not already account for – the specific number of teeth involved in an AOB.

Lastly, the POSI does require an intraoral frontal photo with minimal angular deviation to be of diagnostic quality, and it can be time-consuming to obtain the exact deviation angle of a photo before deciding whether a retake is necessary. The angulation validity threshold using the clinical crowns of the first molars, mentioned previously, is only a rough guide. It must be noted that there are many factors that could necessitate alterations to the threshold guide in the case of abnormalities relating to the first molars. For example, a supraerupted or ankylosed first molar, or a large anteroposterior discrepancy could affect the photographed vertical position of the first molar and thus affect the usefulness for the angulation validity threshold. In the future to avoid issues with angled photos, AOB studies could use intraoral scans instead of traditional photographs. With scans, it could be ensured that the POSI is scored from the frontal view, exactly parallel to the occlusal plane. This could be easily implemented in today's digital age, as more practitioners are including intraoral scans in pre- and post-treatment records.

CONCLUSIONS

1. The results of this study demonstrate that the Photographic Open Bite Severity Index (POSI) is a valid and reliable measure of estimated severity and anticipated treatment difficulty of anterior open bite (AOB), when scored on an intraoral frontal photograph captured from ≤10 degrees above or below the posterior occlusal plane.
2. For the most accurate and reliable POSI scoring, the following conventions should be followed:
    a. When looking for teeth not in "overlap," no portion of the incisal edge may overlap or appear to "touch" the opposing teeth.
    b. For POSI 1-4, only consider the incisors. For POSI 5/6 assume that the tooth pairs with/opposes its matching opposite-arch counterpart.
        i. Visualize an imaginary plane to see if there is "overlap" when the teeth are not photographed to be directly on top of one another.
3. Most POSI scoring errors appeared to be related to misidentifying the opposing tooth/teeth to be scored. Future examiners should be cautioned that diligence is necessary to identify correct teeth to score, especially if a patient presents with a malocclusion characterized by an anteroposterior dentoskeletal discrepancy.
4. While the POSI alone, estimated clinical OB alone, POSI + estimated clinical OB, and POSI + ceph-measured OB were all factors that were able to account for a large portion of the variability in AOB severity and treatment difficulty, the combination of POSI + estimated clinical OB (the COSI) was shown to account for the highest percentage of the variation in AOB severity and treatment difficulty.

a. The Combined Open Bite Severity Index (COSI = 1.55 + 0.50*POSI + 0.39*OB, OB: mm of open bite) accounts for 64% of the variation in AOB severity.

5. Because the ceph-measured OB alone demonstrates the weakest correlation with severity and treatment difficulty of AOB, it is preferable to use the POSI, the COSI, or the millimetric measure of AOB at the position of greatest opening instead as measures of OB severity.
6. The POSI and the COSI can both be utilized as a standardized measure in clinical research, e.g., as an outcome measure in meta-analyses., or even as a standardized measure for open bite in determining treatment need in public health or insurance capacities.

LITERATURE CITED

Abdelkarim A, Jerrold L. 2018. Clinical considerations and potential liability associated with the use of ionizing radiation in orthodontics. American journal of orthodontics and dentofacial orthopedics. 154(1):15-25.

Agbaje H, Osiatuma V, Fadeju A, Kolawole K, Otuyemi O. 2012. Anterior open bite: A review of epidemiology, aetiology and management. West African Journal of Orthodontics. 1(1):5-12.

Altaye M, Dormer A, Klar N. 2001. Inference procedures for assessing interobserver agreement among multiple raters. Biometrics. 57(2):584-588.

Bailey LJ, Haltiwanger LH, Blakey GH, Proffit WR. 2001. Who seeks surgical-orthodontic treatment: A current review. The International journal of adult orthodontics and orthognathic surgery. 16(4):280-292.

Brook PH, Shaw WC. 1989. The development of an index of orthodontic treatment priority. European journal of orthodontics. 11(3):309-320.

Cangialosi TJ, Riolo ML, Owens SE, Dykhouse VJ, Moffitt AH, Grubb JE, Greco PM, English JD, James RD. 2004. The abo discrepancy index: A measure of case complexity. American journal of orthodontics and dentofacial orthopedics. 125(3):270-278.

Carlos JP. 1970. Evaluation of indices of malocclusion. International dental journal. 20(4):606-617.

Choi KW, Ko H-C, Todoki LS, Finkleman SA, Khosravi R, Wang H-F, Funkhouser E, Baltuck C, Raj V, Allareddy V et al. 2018. The national dental practice-based research network adult anterior open bite study: A description of the practitioners and patients. The Angle orthodontist. 88(6):675-683.

Draker HL. 1960. Handicapping labio-lingual deviations: A proposed index for public health purposes. American journal of orthodontics. 46(4):295-305.

Fields HW, Larson BE, Sarver DM, Proffit WR. 2019. Contemporary orthodontics. Philadelphia, PA: Philadelphia, PA : Elsevier.

Fleiss JL. 1981. Statistical methods for rates and proportions. New York: Wiley.

Greenlee GM, Huang GJ, Chen SS, Chen J, Koepsell T, Hujoel P. 2011. Stability of treatment for anterior open-bite malocclusion: A meta-analysis. Am J Orthod Dentofacial Orthop. 139(2):154-169.

Huang G, Baltuck C, Funkhouser E, Wang H-F, Todoki L, Finkleman S, Shapiro P, Khosravi R, Ko H-C, Greenlee G et al. 2019. The national dental practice-based research network adult anterior open bite study: Treatment recommendations and their association with patient and practitioner characteristics. American journal of orthodontics and dentofacial orthopedics. 156(3):312-325.

Kelly JE, Sanchez MJ, Van Kirk LE. 1973. An assessment of the occlusion of the teeth of children 6-11 years, united states.

Kendall MG. 1970. Rank correlation methods. London: London : Griffin.

Lentini-Oliveira DA, Carvalho FR, Rodrigues CG, Ye Q, Prado LB, Prado GF, Hu R. 2014. Orthodontic and orthopaedic treatment for anterior open bite in children. Cochrane Database Syst Rev. (9):CD005515.

Little RM. 1975. The irregularity index: A quantitative score of mandibular anterior alignment. American journal of orthodontics. 68(5):554-563.

Liu S, Oh H, Chambers DW, Baumrind S, Xu T. 2017. Validity of the american board of orthodontics discrepancy index and the peer assessment rating index for comprehensive evaluation of malocclusion severity. Orthodontics & craniofacial research. 20(3):140-145.

Maciel CTV, Leite ICG. 2005. Etiological aspects of anterior open bite and its implications to the oral functions. Pró-fono : revista de atualização científica. 17(3):293-302.

Parker WS. 1998. The hld (calmod) index and the index question. American journal of orthodontics and dentofacial orthopedics. 114(2):134-141.

Proffit WR, Fields JHW, Moray LJ. 1998. Prevalence of malocclusion and orthodontic treatment need in the united states: Estimates from the nhanes iii survey. The International journal of adult orthodontics and orthognathic surgery. 13(2):97-106.

Richmond S, Daniels CP. 1998. International comparisons of professional assessments in orthodontics: Part 1—treatment need. American journal of orthodontics and dentofacial orthopedics. 113(2):180-185.

Richmond S, Shaw WC, Roberts CT, Andrews M. 1992. The par index (peer assessment rating): Methods to determine outcome of orthodontic treatment in terms of improvement and standards. European journal of orthodontics. 14(3):180-187.

Shaw WC, Addy M, Ray C. 1980. Dental and social effects of malocclusion and effectivenessof orthodontic treatment: A review. Community dentistry and oral epidemiology. 8(1):36-45.

Shaw WC, Richmond S, O'Brien KD, Brook P, Stephens CD. 1991. Quality control in orthodontics: Indices of treatment need and treatment standards. British dental journal. 170(3):107-112.

Team RC. 2013. R: A language and environment for statistical computing.

Tsang WM, Cheung LK, Samman N. 1997. Cephalometric parameters affecting severity of anterior open bite. International journal of oral and maxillofacial surgery. 26(5):321-326.

APPENDICES

APPENDIX 1. OCCLUSAL PLANE DEVIATION ANGLE ESTIMATE RELIABILITY

Measure	1st Mean (SD)	2nd Mean (SD)	Diff. (SD) [95% CI]	ICC (95% CI)	D. Error (Min, Max)
Angle estimate	1.6 (3.9)	1.9 (4.2)	-0.3 (0.5) [-0.5, -0.1]	0.99 (0.98, 1.00)	0.4 (0.0, 1.0)

APPENDIX 2. SEVERITY SURVEY INTER-EXAMINER

Examiners	Intraclass correlation coefficient (ICC)	95% CI
All examiners	0.76	0.69 to 0.82
Without examiner #1	0.74	0.67 to 0.81
Without examiner #2	0.73	0.65 to 0.80
Without examiner #3	0.79	0.73 to 0.85
Without examiner #4	0.76	0.69 to 0.82
Without examiner #5	0.77	0.70 to 0.83

APPENDIX 3. AOB SEVERITY SURVEY RESULTS CHARACTERISTICS

		Examiner				
Characteristic	Overall, N = 420	1, N = 84	2, N = 84	3, N = 84	4, N = 84	5, N = 84
Severity						
Median (IQR)	4.5 (3.0, 6.0)	5.0 (4.0, 5.0)	4.5 (3.0, 6.0)	4.0 (3.0, 5.0)	5.0 (3.8, 6.0)	5.0 (3.0, 6.2)
Range	1.0, 10.0	1.0, 10.0	1.0, 9.0	2.0, 9.0	2.0, 9.0	1.0, 10.0
Mean (SD)	4.6 (1.9)	4.7 (1.7)	4.5 (1.9)	4.0 (1.6)	4.7 (1.7)	5.0 (2.3)

AOB severity is a ranking and should be summarized by the median and IQR (25th and 75th percentiles). The mean and standard deviation (SD) are reported for reference.

The Friedman test was used to test for differences in the severity scores amongst the 5 examiners (Friedman test, p-value = <.00001).

APPENDIX 4. SEVERITY SURVEY INTRA-EXAMINER ICC

Examiner	Intraclass correlation coefficient (ICC)	95% CI
#1	0.89	0.73 to 0.96
#2	0.97	0.93 to 0.99
#3	0.94	0.84 to 0.98
#4	0.90	0.75 to 0.96
#5	0.93	0.82 to 0.97

APPENDIX 5. AOB TREATMENT DIFFICULTY SURVEY INTER-EXAMINER ICC

Examiners	Intraclass correlation coefficient (ICC)	95% CI
All examiners	0.67	0.58 to 0.75
Without examiner #1	0.64	0.54 to 0.72
Without examiner #2	0.61	0.51 to 0.71
Without examiner #3	0.68	0.59 to 0.76
Without examiner #4	0.76	0.68 to 0.82
Without examiner #5	0.64	0.55 to 0.73

APPENDIX 6. AOB TREATMENT DIFFICULTY SURVEY RESULTS CHARACTERISTICS

		Examiner				
Characteristic	Overall, N = 420	1, N = 84	2, N = 84	3, N = 84	4, N = 84	5, N = 84
Treatment Difficulty						
Median (IQR)	5.0 (3.0, 6.0)	5.0 (4.0, 6.0)	4.0 (3.0, 6.0)	4.0 (3.0, 5.0)	6.0 (5.0, 7.0)	4.0 (3.0, 6.0)
Range	1.0, 10.0	2.0, 9.0	1.0, 9.0	1.0, 9.0	2.0, 9.0	1.0, 10.0
Mean (SD)	4.7 (1.9)	4.9 (1.6)	4.5 (2.0)	4.1 (1.7)	5.8 (1.6)	4.4 (2.3)

AOB treatment difficulty is a ranking and should be summarized by the median and IQR (25th and 75th percentiles). The mean and standard deviation (SD) are reported for reference.

The Friedman test was used to test for differences in the AOB treatment difficulty scores amongst the 5 examiners (Friedman test, p-value = <.00001).

APPENDIX 7. AOB TREATMENT DIFFICULTY INTRA-EXAMINER ICC

Examiner	Intraclass correlation coefficient (ICC)	95% CI
#1	0.92	0.80 to 0.97
#2	0.94	0.55 to 0.98
#3	0.95	0.56 to 0.98
#4	0.91	0.34 to 0.98
#5	0.78	0.44 to 0.94

APPENDIX 8. GOLD STANDARD POSI CORRELATIONS

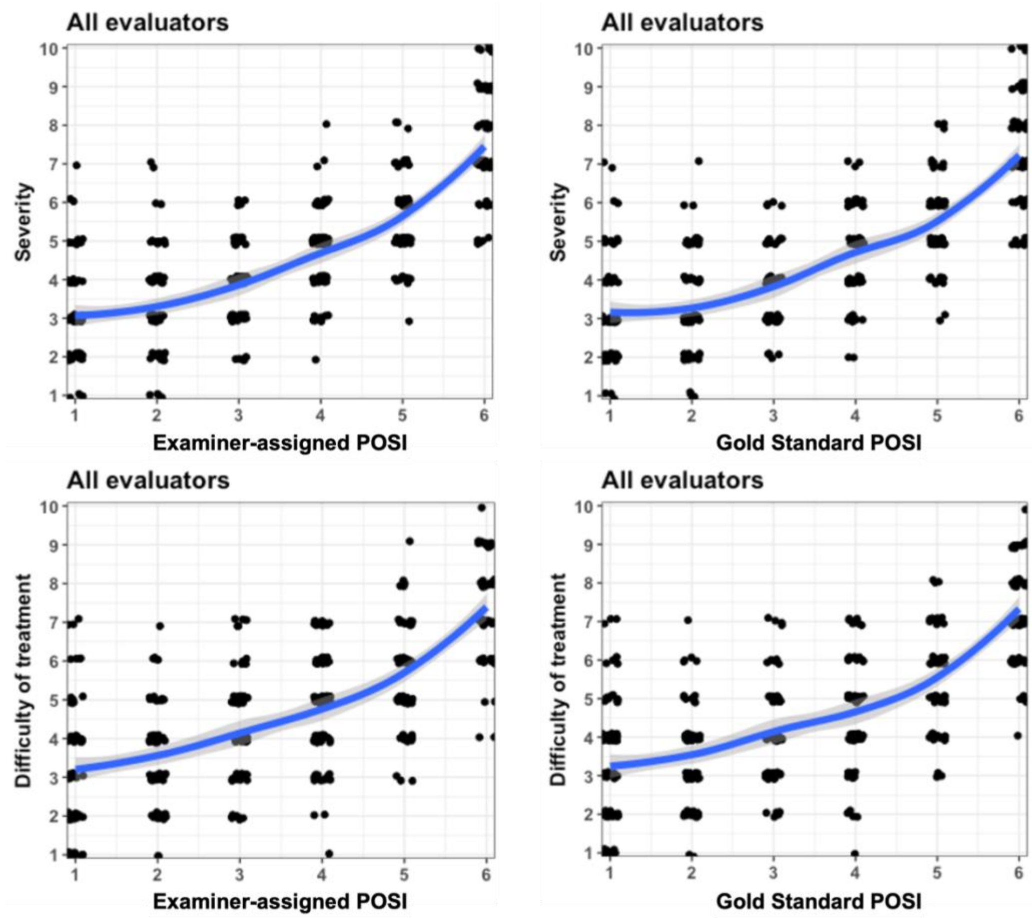
Correlation between gold standard POSI and AOB severity

Examiners	Spearman rank correlation	95% CI
All examiners	0.71	0.66 to 0.75
Without examiner #1	0.72	0.60 to 0.81
Without examiner #2	0.77	0.67 to 0.85
Without examiner #3	0.75	0.63 to 0.83
Without examiner #4	0.69	0.55 to 0.79
Without examiner #5	0.72	0.59 to 0.81

Correlation between gold standard POSI and AOB treatment difficulty

Examiners	Spearman rank correlation	95% CI
All examiners	0.66	0.61 to 0.71
Without examiner #1	0.77	0.67 to 0.85
Without examiner #2	0.81	0.72 to 0.87
Without examiner #3	0.70	0.57 to 0.79
Without examiner #4	0.62	0.47 to 0.74
Without examiner #5	0.70	0.57 to 0.80

APPENDIX 9. SCATTER PLOTS OF POSI VS. AOB SEVERITY OR TREATMENT DIFFICULTY



Appendix 10. Examiner-Assigned POSI Correlations

Correlation between examiner-assigned POSI and AOB severity

Examiners	Spearman rank correlation	95% CI
All examiners	0.73	0.68 to 0.77
Without examiner #1	0.74	0.62 to 0.82
Without examiner #2	0.77	0.66 to 0.84
Without examiner #3	0.75	0.63 to 0.83
Without examiner #4	0.68	0.55 to 0.78
Without examiner #5	0.73	0.62 to 0.82

Correlation between examiner-assigned POSI and AOB treatment difficulty

Examiners	Spearman rank correlation	95% CI
All examiners	0.66	0.61 to 0.71
Without examiner #1	0.76	0.66 to 0.84
Without examiner #2	0.78	0.68 to 0.85
Without examiner #3	0.69	0.56 to 0.79
Without examiner #4	0.62	0.47 to 0.74
Without examiner #5	0.71	0.58 to 0.80

Appendix 11. Correlation between Severity and Treatment Difficulty

Examiners	Spearman rank correlation	95% CI
All examiners	0.76	0.71 to 0.79
Without examiner #1	0.81	0.72 to 0.87
Without examiner #2	0.84	0.77 to 0.90
Without examiner #3	0.85	0.77 to 0.90
Without examiner #4	0.71	0.58 to 0.80
Without examiner #5	0.76	0.65 to 0.84

Appendix 12. Estimated Open Bite Reliability

Measure	1st Mean (SD)	2nd Mean (SD)	Diff. (SD) [95% CI)]	ICC (95% CI)	D. Error (Min, Max)
Overbite estimate	-3.6 (2.5)	-3.5 (2.2)	-0.1 (0.4) [-0.3, 0.1]	0.98 (0.96, 0.99)	0.3 (0.0, 1.0)

Appendix 13. Estimated Clinical Open Bite Validity Test 1

Estimated Ceph OB: an estimation of the vertical opening in millimeters at the most prominent central incisors (where a lateral cephalogram measures OB) obtained by viewing the intraoral frontal photo

Measure	Estimated Ceph OB (SD)	Ceph-measured (SD)	Diff. (SD) [95% CI]	ICC (95% CI)	D. Error (Min, Max)
Estimated Ceph OB vs. Ceph-measured OB	-2.4 (1.8)	-3.2 (2.5)	0.8 (1.3) [0.0, 1.5]	0.78 (0.46, 0.92)	1.0 (0.0, 3.5)

Paired t-test for testing if the averages are the same (or difference), p-value = 0.046

Correlation between the two measures = 0.86 (95% CI, 0.61 to 0.95).

Appendix 14. Estimated Clinical Open Bite Validity Test 2

Ratio: digital measurement from photo of the vertical opening at the most prominent central incisors (where a ceph measures OB) divided by the greatest vertical opening of all anterior teeth (where estimated clinical OB was acquired).

Ratio-calculated OB = Ceph-measured OB/Ratio

Measure	Estimated clinical OB Mean (SD)	Ratio-calculated OB Mean (SD)	Diff. (SD) [95% CI]	ICC (95% CI)	D. Error (Min, Max)
Estimated clinical OB vs. Ratio-calculated OB	-3.8 (2.4)	-4.7 (4.1)	0.9 (2.4) [-0.5, 2.4]	0.72 (0.33, 0.90)	1.8 (0.0, 5.7)

Paired t-test for testing if the averages are the same (or difference), p-value = 0.17

Correlation between the two measures = 0.84 (95% CI, 0.61 to 0.95).

APPENDIX 15. ANGULATION STUDY MODEL CHARACTERISTICS

True POSI	Open bite (mm)	Overjet (mm)
1	1.2	1.9
2	1.8	3.3
3	3.0	4.0
4	1.6	2.0
5	4.4	2.3
6	3.6	2.6

Open bite was measured at most severe location, Overjet was measured at the location where the open bite was measured.