



Daniel A. Paesani

# Reliability of multiple-degree incisal/occlusal tooth wear assessment on dental casts: Findings from a five-examiner investigation and related clinical implications

Daniel A. Paesani, DDS<sup>1</sup>/Luca Guarda-Nardini, MD, DDS<sup>2</sup>/Carlota Gelos, DDS<sup>3</sup>/Luigi Salmaso, PhD<sup>4</sup>/  
 Daniele Manfredini, DDS, PhD<sup>5</sup>

**Objective:** The aim was to answer the clinical research question: is incisal/occlusal tooth wear assessment on dental casts performed by five professionals with expertise in different fields of dentistry reliable? **Method and Materials:** Five examiners with different fields of expertise in the dental profession assessed tooth wear on dental casts of 45 subjects, based on a six-degree rating of incisal/occlusal wear. After a calibration meeting, the examiners evaluated the casts individually and various issues concerning interexaminer agreement and reliability were assessed. **Results:** A total of 872 teeth were evaluated. The five examiners agreed only for the rating of 6.6% of the teeth. The teeth with the highest percentage of agreement were the premolars. Pairwise comparison of the assessments of the examiners #1 (bruxism expert), #2 (orthodontist), #3 (temporomandibular disorder [TMD] and occlusion expert), #4 (dental nurse) showed fair to moderate

agreement, with  $\kappa$ -values ranging from 0.306 to 0.577, whilst the examiner #5 (lab technician) achieved low interexaminer reliability values with all the other four examiners. **Conclusion:** The interexaminer reliability of tooth wear assessment on dental casts performed by five professionals with expertise in different fields of dentistry is highly variable. General practitioners should keep in mind that consensus decisions by the examiners and assessment by raters belonging to the same dental discipline are recommended strategies to increase the reliability of tooth wear evaluation in the clinical setting. **Clinical significance:** This investigation adds to the literature suggesting that, in a clinical setting, a single examiner's assessment of tooth wear on dental casts does not have optimal reliability and that it may be source of internal validity problems in the research setting. (*Quintessence Int* 201#;##:1-6; doi: ##.###/j.qi.a#####)

**Key words:** dental cast, interexaminer reliability, tooth wear

The assessment of tooth wear is a critical issue involving multiple disciplines of the dental profession. Indeed, getting a deeper insight to the etiology (eg,

bruxism, age-related wear, influence of chemical, dietary, and/or other mechanical factors) and clinical consequences (eg, dentinal hypersensitivity, temporomandibular disorders) of tooth wear as well as its prevention and restoration strategies is fundamental for all the dental disciplines dealing with the management of dental occlusion.<sup>1-3</sup> Implementing the diagnosis and rating of tooth wear are key factors for designing methodologically sound studies on the above issues and drawing implications for everyday clinical practice.

<sup>1</sup> Professor, School of Dentistry, University of Salvador/AOA, Buenos Aires, Argentina.

<sup>2</sup> Head, TMD Clinic, Department of Maxillofacial Surgery, University of Padova, Italy.

<sup>3</sup> Private practice, Cordoba, Argentina.

<sup>4</sup> Professor, Department of Management and Engineering, University of Padova, Italy.

<sup>5</sup> Assistant Professor, TMD Clinic, Department of Maxillofacial Surgery, University of Padova, Italy.

**Correspondence:** Dr Daniele Manfredini, Viale XX Settembre 298, 54033 Marina di Carrara (MS), Italy. Email: daniele.manfredini@tin.it

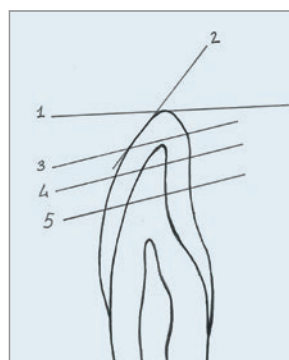
Tooth wear can be assessed instrumentally, clinically, or by the evaluation of dental casts. Over the years, several strategies have been proposed to rate it for different purposes in the research as well as the clinical setting, but it seems that there is not one ideal evaluation method or index.<sup>4</sup>

Also, despite the number of different indexes that are currently available to assess tooth wear, reliability data were provided only in a few papers. Reliability analysis of tooth wear assessment on dental casts showed, in general, a fair to good agreement between two examiners.<sup>5</sup> Nonetheless, it must be pointed out that data on multiple-examiner reliability are lacking and that most papers on various tooth wear topics were based on an assessment performed by a single examiner. This may represent a potential flaw for the internal validity of findings. Besides, if one considers the multidisciplinary interest in the diagnosis of tooth wear, the risk that its assessment is influenced by the discipline of the dental professional performing the evaluation cannot be underestimated.

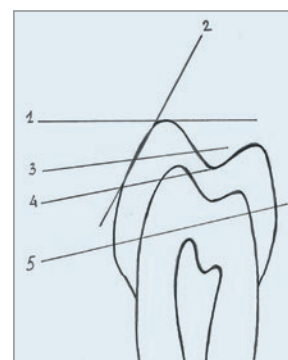
Based on this premise, this investigation was specifically designed to answer the clinical research question: is incisal/occlusal tooth wear assessment on dental casts performed by five professionals with expertise in different fields of dentistry reliable?

## METHOD AND MATERIALS

Forty-five patients (mean age  $40.3 \pm 8.2$  years) attending a dental clinic were recruited and gave their informed consent to take part in the study. For all patients, impressions of both dental arches with irreversible hydrocolloid (Alginoplast, Heraeus Kulzer) were made. Five dental professionals with different roles and fields of expertise assessed tooth wear on dental casts: a bruxism and temporomandibular disease (TMD) practitioner with expertise in tooth wear assessment (DAP, examiner #1), an orthodontist with expertise in functional assessment of jaw movements (CG, #2), an orofacial pain practitioner with expertise in TMD and occlusion (DM, #3), a dental nurse with expertise in tooth observation for clinical records during



**Fig 1** Rating of tooth wear in the anterior and canine teeth.



**Fig 2** Rating of tooth wear in the premolar and molar teeth.

gnathological assessment (TB, #4), and a lab technician with expertise in gnathology and fixed prosthodontics (ML, #5). The study consisted of a calibration meeting and an investigation session performed in two successive days. During the calibration meeting, the five examiners evaluated dental casts of five randomly selected patients to rate tooth wear by consensus and set the parameters to be used in the investigation. During the investigation session, the examiners independently rated the wear of each tooth on the remaining 80 dental casts. The sequence of examination was based on a rotation of dental casts among the examiners. Parameters chosen for the assessment of tooth wear were retrieved from literature suggestions<sup>6,7</sup> and were established as follows on the basis of a possible six-degree rating (Figs 1 and 2):

- 0 = no tooth wear
- 1 = slight wear on the top of the cusps or incisal tips
- 2 = noticeable wear in the form of flattening with respect to the normal contour of the cusps or tips
- 3 = marked flattening of the cusps or tips
- 4 = total loss of cuspal or tips contour and moderate dentinal exposure
- 5 = severe tooth wear with marked dentinal exposure.

Teeth with ceramic crowns, gross restorations of the occluding surfaces, decayed and/or broken teeth as well as the wisdom teeth were excluded from the analysis. The percentage of teeth receiving the same



**Table 1** Percentage of different ratings by the five examiners with respect to the total tooth sample (N = 872)

Tooth wear rating	#1	#2	#3	#4	#5
0	39.1	21.7	45.9	53.0	5.1
1	7.6	7.7	14.9	23.3	53.8
2	16.2	41.0	17.4	12.6	18.2
3	17.9	6.4	6.1	6.0	19.1
4	15.3	19.7	12.5	3.6	2.3
5	3.9	3.4	3.3	1.5	1.5

**Table 2** Agreement between the examiners: percentage of teeth receiving the same wear score by the specified number of examiners

Level of agreement	No. of examiners in agreement	% of teeth
Full disagreement	1	0.4
	2	30.9
	3	38.1
Partial agreement	4	24.0
	5	6.6

rating by the different examiners (ie, five, four, three, two, or one examiners assigning the same score) was calculated. The percentage of examiners in agreement was also assessed with respect to each tooth position. Pairwise comparisons were performed to assess inter-examiner reliability by means of kappa statistics, calculated with 95% confidence intervals (CIs). The qualitative interpretation of kappa-values was based on the following parameters:<sup>8</sup>

- 0.01 to 0.20 slight agreement
- 0.21 to 0.40 fair agreement
- 0.41 to 0.60 moderate agreement
- 0.61 to 0.80 substantial agreement
- 0.81 to 0.99 almost perfect agreement.

Also, the concordance correlation coefficient was calculated for each pair of examiners, to evaluate the degree to which pairs of observations fall on the 45° line through the origin.<sup>9</sup> Finally, intraclass correlation analysis was performed to assess the influence of the different ratings on the reliability values.<sup>10</sup> All statistical procedures were performed with the software SPSS 19.0 (SPSS).

## RESULTS

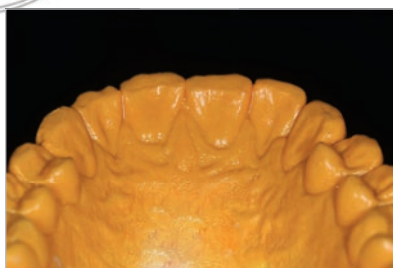
In total, after excluding missing, broken, decayed, restored, and third molar teeth, 872 teeth were eligible for the assessment of wear rating. The different ratings were assigned with variable frequency between the

five examiners. The most frequent ratings were: degree 0 for the examiners #1 (39.1% of the teeth), #3 (45.9%), and #4 (53.0%), degree 1 for the examiner #5 (53.8%), and degree 2 for the examiner #2 (41.0%) (Table 1).

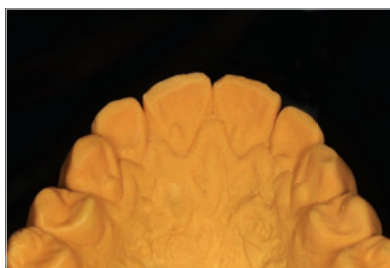
The five examiners were in full agreement (ie, all examiners assigning the same wear score) for the assessment of only 6.6% of the teeth. In almost one third of the teeth a low agreement (ie, agreement by only two examiners or complete disagreement between the examiners) was shown (Table 2). Within the subsample of teeth receiving full agreement between the examiners, 32.1% were rated 0, 26.4% were rated 4, and 24.5% were rated 5. Wear ratings 2 (7.5%), 1 (5.7%), and 3 (3.8%) were less prevalent between the subsample of teeth receiving full agreement, suggesting that the assessment of teeth with no wear or severe wear was easier than that of teeth with intermediate wear scores (Fig 3).

As for the interexaminer agreement with respect to tooth position, the number of examiners agreeing on the evaluation of each specific tooth was calculated. The teeth with the highest percentage of at least four-examiner agreement were the premolars, with a percentage ranging from 39.4% to 63.7%. On the contrary, the teeth with the lowest agreement were the incisors, with a percentage of teeth rated in agreement by two or less examiners ranging from 38.2% to 50% (data not shown) (Figs 4 and 5).

The highest interexaminer reliability was shown by the assessments of the examiners #1 and #3 ( $\kappa = 0.577$ )



**Fig 3** Severe tooth wear conditions can be easily assessed. In this investigation, approximately 50% of the teeth with full agreement between the examiners were rated 4 or 5.



**Fig 4** The interpretation of wear conditions with or without dentinal exposure on the palatal surfaces of the maxillary incisors was the condition with the lowest interexaminer agreement in this investigation.



**Fig 5** Noticeable wear on the top of the cusps in the premolars was the condition with the highest interexaminer agreement in this investigation.

(Table 3), which were also the two examiners showing the highest concordance correlation coefficient ( $r = 0.697$ ) (Table 4). On the contrary, the examiner #5 achieved low interexaminer reliability values with all the other four examiners, with  $\kappa$  values ranging from 0.212 to 0.261. Pairwise comparison of the assessments of the examiners #1, #2, #3, #4 showed fair to moderate agreement, with  $\kappa$ -values ranging from 0.306 to 0.577.

Intraclass correlation coefficient was 0.575, thus suggesting that the variability of the assessment was influenced by the degree of tooth wear.

## DISCUSSION

Measuring tooth wear is an important clinical and scientific requirement for different dental professionals involved in diagnosing and restoring the worn dentition. Over the years, several indexes have been proposed to rate tooth wear at chairside or on dental casts. The evolution of those indexes reflects a progressive introduction of more complex rating strategies.<sup>4</sup> Nonetheless, reliability analysis was performed only in a few investigations, and the majority of papers on various topics related with tooth wear (eg, bruxism, orofacial pain, prosthetic dentistry) were based on a single-examiner evaluation, thus representing a potential flaw for the internal validity of findings. Therefore, a consensus standard of reference for tooth wear assessment is not available.

In general, literature studies reported a lower between-examiner agreement for the assessment on

dental casts with respect to clinically-based tooth wear evaluation.<sup>5</sup> The few studies on dental casts were mainly based on the comparison between ratings from two examiners belonging to the same dental discipline. For instance, an investigation involving six oral surgeons retrieved a fair to moderate reliability as for tooth wear assessment on dental casts,<sup>6</sup> with  $\kappa$ -values of 0.20 to 0.46. Such findings were in line with those reported in an early study on seven examiners adopting a more complicated and time-consuming tooth wear index.<sup>11</sup> Some previous studies on two or three examiners reported reliability in terms of percentage agreement between the examiners, with a range from around 84% to 91%,<sup>12,13</sup> so that a direct comparison between findings from the different studies is not achievable. In addition, more recent studies involving two examiners showed that agreement was dependent on the type of teeth<sup>14</sup> and that the correlation coefficient was higher for the premolars than the other teeth (0.766 vs 0.628 to 0.748).<sup>5</sup> Based on these premises, it emerged that the available knowledge on interrater reliability might be improved by performing an investigation involving multiple examiners belonging to different dental disciplines and having different fields of expertise.

The rating index adopted in this investigation was based on the assessment of tooth wear on the occlusal surfaces and the changes in the cuspal or tips contours. The rationale for using such a relatively simple index was that it may be suitable to assess the interrater reliability between multiple examiners not belonging to the same dental discipline, so reducing the risk for interpretation



**Table 3** Kappa statistic (95% CI) for pairwise assessment of interexaminer reliability

Examiner	#1	#2	#3	#4	#5
#1	NA	0.498 (0.45 – 0.53)	0.577 (0.53 – 0.61)	0.494 (0.44 – 0.53)	0.223 (0.17 – 0.27)
#2	0.498 (0.45 – 0.53)	NA	0.531 (0.49 – 0.57)	0.306 (0.26 – 0.34)	0.212 (0.17 – 0.25)
#3	0.577 (0.53 – 0.61)	0.531 (0.49 – 0.57)	NA	0.353 (0.31 – 0.39)	0.237 (0.19 – 0.27)
#4	0.494 (0.44 – 0.53)	0.306 (0.26 – 0.34)	0.353 (0.31 – 0.39)	NA	0.261 (0.21 – 0.30)
#5	0.223 (0.17 – 0.27)	0.212 (0.17 – 0.25)	0.237 (0.19 – 0.27)	0.261 (0.21 – 0.30)	NA

NA, not applicable

**Table 4** Concordance correlation coefficients (95% CI) for pairwise assessment of interexaminer reliability

Examiner	#1	#2	#3	#4	#5
#1	NA	0.638 (0.60 – 0.67)	0.697 (0.66 – 0.73)	0.634 (0.59 – 0.67)	0.331 (0.27 – 0.38)
#2	0.638 (0.60 – 0.67)	NA	0.672 (0.63 – 0 – 70)	0.412 (0.37 – 0.45)	0.321 (0.26 – 0.37)
#3	0.697 (0.66 – 0.73)	0.672 (0.63 – 0 – 70)	NA	0.469 (0.42 – 0.51)	0.390 (0.34 – 0.43)
#4	0.634 (0.59 – 0.67)	0.412 (0.37 – 0.45)	0.469 (0.42 – 0.51)	NA	0.406 (0.35 – 0.45)
#5	0.331 (0.27 – 0.38)	0.321 (0.26 – 0.37)	0.390 (0.34 – 0.43)	0.406 (0.35 – 0.45)	NA

NA, not applicable

bias. Nonetheless, this study showed that the evaluation of tooth wear was not consistent between the five examiners, with full agreement only in 6.6% of the teeth. It was also shown that the most reliable interexaminer evaluation can be performed on teeth with the extreme wear degrees (ie, without wear or with severe wear), whilst the intermediate conditions of light-to-moderate wear of the cusps or tips were less reliably assessed.

Interestingly, a pattern of recurrent interpretations for each specialist was identified, despite the baseline calibration meeting. In particular, whilst the two examiners with expertise in bruxism/TMD/occlusion (#1 and #3) showed the highest interrater correlation value, thus suggesting that professionals belonging to the same dental discipline may reach higher reliability values, some speculations on the other professionals' interpretation of tooth wear can be made. For instance, the orthodontist (#2) tended to detect light wear of the cusps along the functional trajectories, whilst the lab technician (#5), on average, rated more severe wear, possibly due to his focus on the detection of variations from normal dental anatomy. These potential biases related to an examiner's specific field of expertise must

be addressed with future research, and need to be taken into account when approaching the literature on tooth wear in different dental disciplines.

From a methodologic viewpoint, this investigation was designed to address previous studies' limitations, such as the low number of examiners and casts as well as the lack of interdisciplinary involvement. On the other hand, no information was gathered on intrarater reliability, which is another issue to investigate in future research. It must also be pointed out that severe tooth wear was rare in the analyzed sample; thus, the fact that the frequency of different wear situations may have an effect on the reliability of the results cannot be excluded on the basis of the less than optimal intraclass correlation coefficient. Future studies should be designed to address this issue as well.

From a clinical viewpoint, this study suggests that several factors depending on either the subjective experience of the professional or the wear degree and position of the tooth may influence this kind of evaluation in the clinical setting. This means that some specialist literature on the diagnosis, clinical relevance, and management of tooth wear, such as investigations





dealing with some bruxism-related topics,<sup>15,16</sup> is potentially flawed by the unclear reliability of tooth wear assessment. These findings are especially important for those general practitioners who are less accustomed with research issues, since they suggest that evaluating the degree of tooth wear is a complex task that requires good practical skills and the optimization of technical aspects. In particular, general dentists should be conscious that personal opinions related to the practitioner's own background and expertise may strongly influence the interpretation of dental casts; this may be important in cases where a differential diagnosis should be put between the different causes and etiologies of tooth wear, and makes a multiple-examiner discussion fundamental to improve the validity of the interpretation. Of course, the data here discussed should stand comparison with more sophisticated approaches (eg, scanners, computers) to detect tooth wear. In that sense, whilst there are no doubts that computer-based strategies may be helpful to increase knowledge on the topic, it must also be remarked that they may not be feasible in everyday clinical practice from a cost-benefit viewpoint.

In summary, this investigation adds to the literature showing that multiple-examiner reliability of tooth wear assessment is far from optimal, even if the examiners have been previously calibrated. It also suggests that factors such as the type/position of the tooth (ie, anterior teeth, premolars, molars), the field of expertise or discipline of the examiner, and the wear severity may strongly influence any reliability assessment of tooth wear on dental casts.

## CONCLUSION

The answer to the clinical research question underlying this investigation is that the interexaminer reliability of tooth wear assessment on dental casts performed by five professionals with expertise in different fields of dentistry is highly variable, likely depending on several factors (eg, wear degree, position of the teeth, field of interest of the examiner) that were investigated in this study. Based on these findings, as an important clinical

remark, it can be recommended that investigations on this issue are based on consensus decisions by the examiners to avoid internal validity problems, and that tooth wear is assessed by raters belonging to the same dental discipline.

## ACKNOWLEDGMENTS

The authors of this manuscript thank Mrs Teresa Bertozzi, dental nurse, and Mr Massimo Labori, lab technician, for their precious cooperation during the assessments of dental casts.

## REFERENCES

1. Manfredini D, Lobbezoo F. Relationship between bruxism and temporomandibular disorders: a systematic review of literature from 1998 to 2008. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2010;109:e26–50.
2. Paesani DA. Dental Erosion. In: Paesani DA (ed). *Bruxism: theory and practice*. Berlin: Quintessence Publishing 2010:149–186.
3. Johansson A, Johansson AK, Omar R, Carlsson GE. Rehabilitation of the worn dentition. *J Oral Rehabil* 2008;35:548–566.
4. Bardsley PF. The evolution of tooth wear indices. *Clin Oral Investig* 2008;12(Suppl 1):S15–S19.
5. Wetselaar P, Lobbezoo F, Koutris M, Visscher CM, Naeije M. Reliability of an occlusal and nonocclusal tooth wear grading system: clinical use versus dental cast assessment. *Int J Prosthodont* 2009;22:388–390.
6. Hooper SM, Meredith N, Jagger DC. The development of a new index for measurement of incisal/occlusal tooth wear. *J Oral Rehabil* 2004;31:206–212.
7. Seligman DA, Pullinger AG. Dental attrition models predicting temporomandibular joint disease or masticatory muscle pain versus asymptomatic controls. *J Oral Rehabil* 2006;33:789–799.
8. Fleiss JL. The measurement of interrater agreement. In: Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. New York: John Wiley & Sons, 1981:212–236.
9. Lin LK. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255–268.
10. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428.
11. Smith BG, Knight JK. An index for measuring the wear of teeth. *Br Dent J* 1984;156:435–438.
12. Øilo G, Dahl BL, Hatle G, Gad AL. An index for evaluating wear of teeth. *Acta Odontol Scand* 1987;45:361–365.
13. Johansson A, Haraldson T, Omar R, Kiliaridis S, Carlsson GE. A system for assessing the severity and progression of occlusal tooth wear. *J Oral Rehabil* 1993;20:125–131.
14. Pergamalian A, Rudy TE, Zaki HS, Greco CM. The association between wear facets, bruxism, and severity of facial pain in patients with temporomandibular disorders. *J Prosthet Dent* 2003;90:194–200.
15. Janal MN, Raphael KG, Klausner J, Teaford M. The role of tooth-grinding in the maintenance of myofascial face pain: a test of alternate models. *Pain Med* 2007;8:486–496.
16. John MT, Frank H, Lobbezoo F, Drangsholt M, Dette KE. No association between incisal tooth wear and temporomandibular disorders. *J Prosthet Dent* 2002;87:197–203.