Received: 22 November 2020 | Revised: 21 December 2020 | Accepted: 22 December 2020

DOI: 10.1111/joor.13141

ORIGINAL ARTICLE

JOURNAL OF ORAL REHABILITATION WILEY

A comparative evaluation between the reliability of gypsum casts and digital greyscale intra-oral scans for the scoring of tooth wear using the Tooth Wear Evaluation System (TWES)

Shamir B. Mehta[1,2] | Ewald M. Bronkhorst[1] | Luuk Crins[1] | Marie-Charlotte D. N. J Huysmans[1] | Peter Wetselaar[3] | Bas A. C. Loomans[1]

[1]Department of Dentistry, Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands

[2]Department of Conservative & MI Dentistry, Unit of Distance Learning, King's College London Faculty of Dentistry, Oral & Craniofacial Sciences, Guy's Campus, London, UK

[3]Department of Orofacial Pain and Dysfunction, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam, Amsterdam, The Netherlands

Correspondence
Shamir B. Mehta, King's College London, Faculty of Dentistry, Oral & Craniofacial Sciences, Conservative & MI Dentistry, Unit of Distance Learning, Floor 18, Tower Wing, Guy's Campus, St Thomas's Street, SE1 9RT London, UK.
Email: shamir.mehta@kcl.ac.uk

Abstract

Background: The Tooth Wear Evaluation System (TWES) is a type of tooth wear index. To date, there is the lack of data comparing the reliability of the application of this index on gypsum cast records and digital greyscale intra-oral scan records.

Objectives: A comparative evaluation between the use of gypsum cast records and digital greyscale intra-oral scan records with the reliability of tooth wear scoring using the TWES amongst a group of patients with tooth wear.

Methods: Records for 10 patients with moderate to severe tooth wear (TWES ≥ 2) were randomly selected from a larger clinical trial. TWES grading of the occlusal/incisal, buccal and palatal/lingual surfaces was performed to determine the levels of intra- and interobserver agreement. Intra-observer reproducibility was based on the findings of one examiner only. For the interobserver reproducibility, the findings of two examiners were considered. One set of models/ records were used per patient. Cohen's weighted kappa (κ_W) was used to ascertain agreement between and within the observers. Comparison of agreement was performed using t tests ($P < .05$).

Results: For the scoring of the of the total occlusal/incisal surfaces, the overall levels of intra- and interobserver agreement were significantly higher using the gypsum cast records than with the digital greyscale intra-oral scan records, ($P < .001$) and ($P < .001$), respectively. For the overall buccal surfaces, only a significant difference was found in the intra-observer agreement using gypsum casts, ($P = .013$). For the palatal/lingual surfaces, a significant difference was only reported in the interobserver agreement using gypsum casts, ($P = .043$). At the occlusal/incisal surfaces, grading performed using gypsum casts, culminated in significantly higher TWES scores than with the use of the digital greyscale intra-oral scans ($P < .001$). At the buccal and palatal/lingual surfaces, significantly higher wear scores were obtained using digital greyscale intra-oral scan records ($P < .009$).

Conclusions: The TWES can offer a reliable means for the scoring of wearing occlusal/incisal surfaces using gypsum casts. The reliability offered by digital greyscale intra-oral scans for consecutive scoring was in general, inferior.

KEYWORDS
assessment tools, dental casts, digital casts, grading scales, reliability, tooth wear, tooth wear evaluation system (TWES)

1 | BACKGROUND

In 2018, an estimated mean global prevalence of erosive tooth wear in permanent teeth between 20% and 45% was described.[1] Tooth wear can result in a variety of dentofacially related symptoms, to include, aesthetic impairment, sensitivity, pain, discomfort and/ or functional problems.[2,3] More severe forms of tooth wear may also have an adverse impact on a patient's quality of life.[4-6]

Restorative intervention is sometimes prescribed for patients with tooth wear.[3] However, treatment (with a direct resin composite technique, or indirect techniques) may prove to be costly and complex.[7] There may also be some ambiguity with the optimal timing for restorative intervention.[3,8] Whilst counselling and monitoring are advised for all patients with pathological tooth wear, restorative intervention may be indicated when the presenting tooth wear is a clear concern for the patient and/or the clinician, where there may be functional, or aesthetic concerns and/or symptoms of pain, or discomfort.[3] However, definitive dental restorations for tooth wear management should not be prescribed until any active dental pathology has been effectively managed and full patient commitment is available.[3] Where the presenting pathological tooth wear is not progressive and with the lack of any further concerns, restorative intervention may not be necessary and management with vigilant monitoring and counselling, may be continued.[3]

Determining the most appropriate time to prescribe restorative intervention should also consider the progression of the wear process.[3] The need for pragmatic and reliable means to assess the rate of tooth wear progression (between appointments, as well as between different clinicians) is therefore relevant. Tooth wear assessment is most frequently undertaken by periodic clinical (chairside) assessment; however, photographs, serial (consecutive) dental casts and serial digital 3D data scans may also be used to undertake assessment, each with their own limitations.[3,9]

A plethora of tooth wear indices have been introduced for the scoring of the severity of the tooth wear present,[10-15] but the universal acceptance of a grading scale for erosive tooth wear in general dental practice, is lacking.[16] A clinical tooth wear index should ideally offer the potential to undertake scoring using indirect methods such as intra-oral photographs, traditional gypsum dental casts and on digital intra-oral scans,[17] thereby enabling some extra-oral assessment. This may be particularly beneficial when the available clinical chairside time may be constrained.

The Tooth Wear Evaluation System (TWES) is a modular clinical guideline that can be used for the assessment of tooth wear and to assist with diagnosis and patient management.[14,15,18] The TWES was revised in 2020 and a new taxonomy was proposed—TWES 2.0.[18] The TWES index in general includes the application of an 8-point occlusal/incisal ordinal grading scale and a 3-point non-occlusal/ non-incisal grading scale for the scoring of the respective surfaces. The TWES has been reported to offer adequate levels of reliability with tooth wear grading when applied clinically, as well as when using dental cast records.[19] Furthermore, when undertaking occlusal/ incisal surface grading using dental casts and intra-oral photographic records, the TWES has been described to offer the necessary sensitivity to enable the detection of changes in the pattern on tooth wear on a sequential basis and, thereby, help monitor disease progression.[17,20] The aim of this study was to undertake a comparative evaluation between the use of gypsum casts and digital greyscale (black-white) intra-oral scan records with the reliability of grading tooth wear using the TWES, applied to patient records that were demonstrative of moderate to severe forms of tooth wear.

2 | MATERIALS & METHODS

2.1 | Tooth wear evaluation system

The Tooth Wear Evaluation (TWES) was used as the grading system in this investigation.[14,15,18] For the scoring of the occlusal and incisal surfaces, an 8-point ordinal scale was used. The grades defined as grade 0 = no (visible) wear; grade 1a = minimal wear of cusps or incisal tips, within the enamel; grade 1b = facets parallel to the normal planes of contour, within the enamel; grade 1c = noticeable flattening of cusps or incisal edges, within the enamel; grade 2 = wear with dentine exposure and loss of clinical crown height <1/3; grade 3a = wear with dentine exposure and loss of clinical crown height 1/3-1/2; grade 3b = wear with dentine exposure and loss of clinical crown height >1/2-2/3; and grade 4 = wear with dentine exposure and loss of clinical crown height of >2/3. For scoring at the non-occlusal/non-incisal surfaces, a 3-point ordinal scale was applied. The grades are described as grade 0 = no (visible) wear; grade 1 = wear confined to the enamel; and grade 2 = wear into the dentine. The scope of this study is based on the TWES. It did not include the extensions from the TWES 2.0, as data collection commenced prior to the introduction of the updated taxonomy.

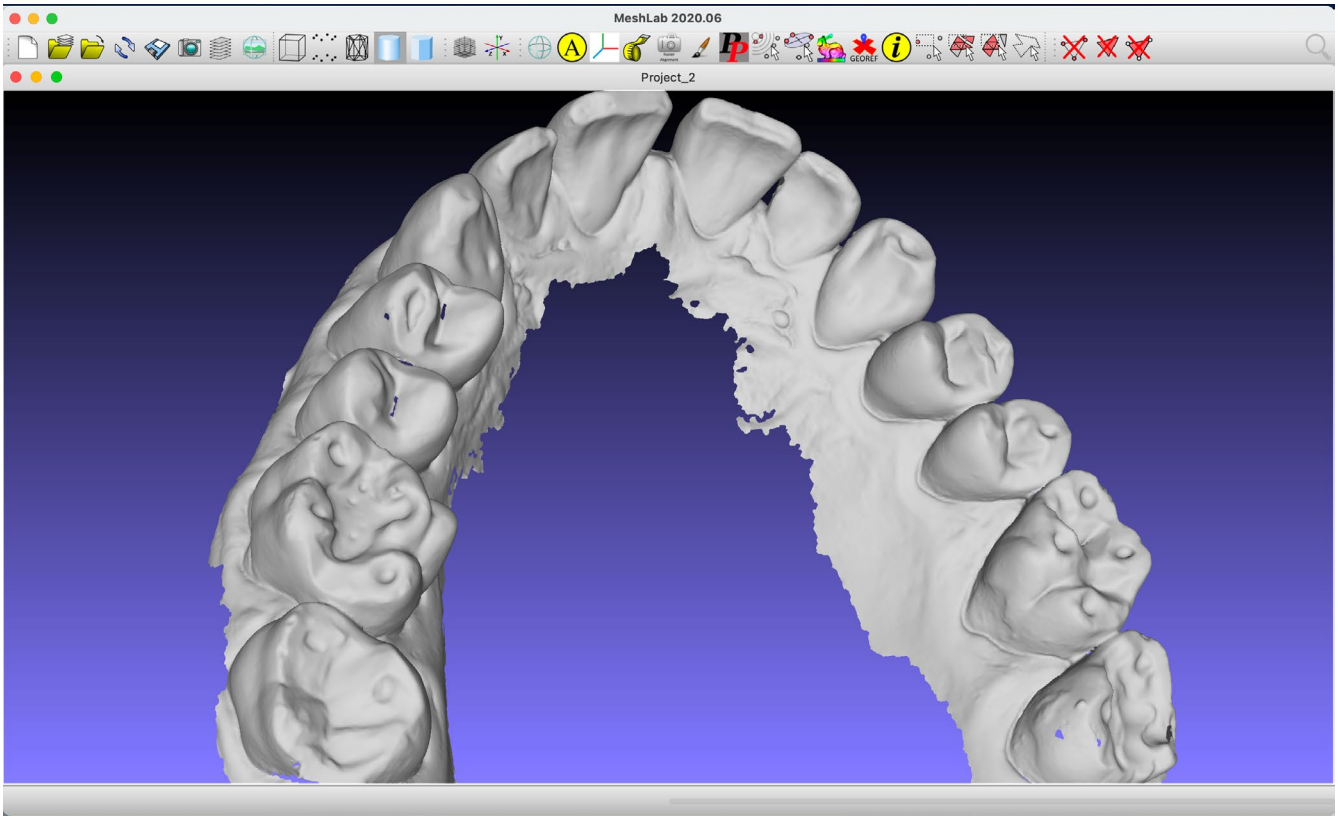

FIGURE 1 The use of Meshlab to score intra-oral scans with the TWES [Colour figure can be viewed at wileyonlinelibrary.com]

2.2 | Subjects

The current investigation is part of a larger clinical trial on the management of erosive tooth wear, the Radboud Tooth Wear Project, in which 200 patients are included.[21] The records of ten patients were randomly selected for the present study. Inclusion criteria were the presence of moderate to severe tooth wear, with at least one score of TWES ≥ 2. The records applied in this investigation were limited to gypsum casts and 3D (three dimensional) digital intra-oral scans. The study was carried out in accordance with the Declaration of Helsinki for research involving humans and ethical approval was obtained (ABR codes: NL31401.091.10, NL30346.091.10 and NL31371.091.10). All patients agreed to participate in the research project, and written informed consent was attained prior to entering the Radboud Tooth Wear Project.

The baseline dental condition of each participant had been fully documented, and full-arch gypsum casts of the upper and lower dental arches were fabricated. Dental impressions were taken using a vinyl polysiloxane impression material, (Ivoclar Virtual 380, Ivoclar Vivodent, Liechtenstein, Europe) comprising two consistencies, a Heavy body and Monophase applied in a single stage. The impressions were cast in Type III dental stone (SLR Dental GmbH, Germany) within 24 hours, according to the manufacturer's instructions. A yellow-coloured dental stone material was used. During the same appointment, digital intra-oral scans were obtained using the LAVA COS Intraoral Scanner (3M). Both the digital and dental impressions were captured by the same trained operator. The scanning procedure was undertaken in accordance with the manufacturer's instructions. Scans were made with the patient in a supine position, a latex-free lip and cheek retractor was applied, Optragate (Ivoclar Vivadent, Liechtenstein), teeth were rinsed, air-dried and lightly powdered with titanium dioxide. The LAVA COS scanner was used to capture the digital impression, including the bite registration scan. The scans were digitally stored in the web-based platform, Casemanager (3M). The 3D models of the scans ('digital intra-oral scans/ digital models') were amenable to downloading from this platform and these open STL files could be easily imported into the free-software, MeshLab (www.meshlab.net). Figure 1 is a representation of the MeshLab user interface.

2.3 | Scoring and the intra- and interobserver agreement

In advance of this study, Observer 1 (O1), a final-year undergraduate dental student, was trained and calibrated over the course of two training sessions with the use of the TWES by Observer 2 (O2). Observer 2 was an experienced dental practitioner and researcher.

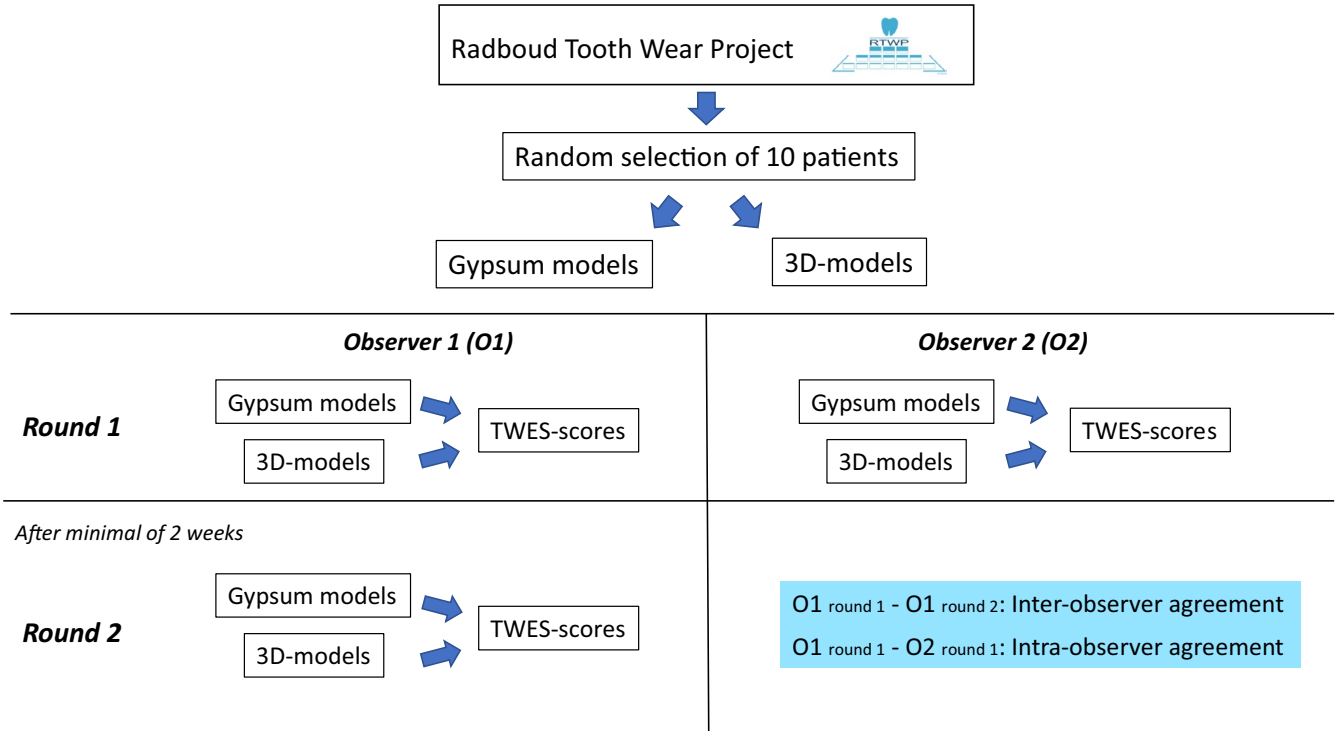

FIGURE 2 Flowchart of assessment protocol: intra- and inter-observer agreement [Colour figure can be viewed at wileyonlinelibrary.com]

The gypsum cast records included in this investigation were scored using the TWES in the same environment and appraised under consistent, standard room lighting conditions. Under same conditions, the digital intra-oral scan records were visualised in greyscale on a computer screen (resolution: 1920x1080) with MeshLab, enabling the assessor to rotate and zoom in on the models. As the output of the 3D models when using the LAVA COS scanner is in greyscale, this formed the rationale for the use of greyscale records in this investigation. The sequence of scoring for all records was, the first quadrant, followed by the second, the third and finally, the fourth. No time limit was set for the evaluations.

Teeth with fixed prosthodontic restorations (eg crowns and bridges), or large intra-coronal restorations were excluded from the analysis. Teeth that were not clearly visible (inclusive of teeth that were unclear on the digital intra-oral scans), or where they were broken/ or damaged on the casts were also excluded from the analysis.

For the intra-observer measurements, the ten sets of gypsum casts and digital intra-oral scan records were scored twice by Observer 1 with a minimum interval of 2 weeks between the consecutive observations. Comparisons were made between the consecutive scores for the full mouth (overall scores), as well as for anterior and posterior areas. Assessments were then undertaken by Observer 2 applying the same protocols; however, for the purpose of evaluating the interobserver agreement, only one round of scoring was performed by O2. To study the interobserver agreement (O1-O2), the gypsum casts and digital greyscale intra-oral scan records for the same ten cases were scored once by both observers, O1 and O2. The observers were blinded to each other's scores and in the case of O1 blinded to the outcomes of their former observations when carrying out the second round of their assessments. In Figure 2, a flow diagram has been provided to summarise the assessment protocol.

To evaluate the effect of the 'type of record' (gypsum models or digital greyscale intra-oral scans) on the scoring with the TWES, the differences in Observer 1's tooth wear scores at each of the surfaces assessed using the gypsum models and digital scan records were determined.

TABLE 1 Conversion of the TWES grades into numerical scores, as applied in this investigation

TWES grade	0	1a	1b	1c	2	3a	3b	4
Numerical score	1	2	3	4	5	6	7	8

Note: TWES grades as per Wetselaar & Lobbezoo, 2016.

2.4 | Statistical analyses

To describe the agreement between the intra- and interobserver scores for the assessments using the gypsum cast records or digital intra-oral scans, Cohen's weighted Kappa (κ_W) was used. In all Kappa analyses, squared weights were applied. Kappa measures were interpreted as follows: <0 as indicating 'no agreement,' 0-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial and 0.81-1 as almost perfect agreement.[22] Scores were presented for the 'overall' (total) occlusal/ incisal surfaces, for the

TABLE 2 Descriptives of tooth wear scores using the TWES at all tooth surfaces, measured on gypsum casts (n = 10)

	Occlusal/incisal grading (8-point scale)								Buccal grading (3-point scale)			Lingual/Palatal grading (3-point scale)		
Tooth type	**0**	**1a**	**1b**	**1c**	**2**	**3a**	**3b**	**4**	**0**	**1**	**2**	**0**	**1**	**2**
Incisors	0	0	0	6	31	17	13	11	14	64	0	14	64	0
Canines	0	0	0	0	16	9	8	7	3	37	0	8	32	0
Premolars	0	0	0	0	51	18	5	2	6	70	0	41	34	0
Molars	0	2	2	3	46	22	4	1	28	0	0	39	40	0
Total	0	2	2	9	144	66	30	21	51	223	0	102	170	0

buccal surfaces and for the palatal/lingual surfaces. Scores were also presented by tooth type, hence, anterior teeth (incisors and canines) and posterior teeth (premolars and molar teeth), irrespective of the arch. Differences in Kappa scores were analysed using t tests and the data expressed as mean values, with confidence intervals, (95% ci) and *P*-values ($P < .05$).

To determine the effect of the type of record on the scoring outcome, the TWES scores of the occlusal/incisal scale (0, 1a, 1b, etc) were converted into numerical scores, ranging from 1 to 8 inclusive. Hence, as seen by Table 1, a TWES outcome of '0' would be scored '1', 1a as '2', 1b as '3' etc For all measurements, the scores of the digital intra-oral scans and the gypsum cast records were compared with a paired *t* test. The mean difference in the TWES scores at the various surfaces were evaluated for the two types of records; a positive score would indicate scoring using a gypsum model would result in a higher TWES score. All analyses were performed using R (version 3.6.1). Weighted kappa values were calculated using the Kappa function of the vcd library (version 1.4-7).

3 | RESULTS

Table 2 provides a combined overview of the TWES scores for all ten patient records at the occlusal/incisal, the buccal and the palatal/lingual surfaces. The patient records showed the presence of significant amounts of tooth wear at all teeth. The majority of the scores at the occlusal/incisal surfaces were between TWES 2 (showing wear with dentine exposure and loss of clinical crown height <1/3) and TWES 3b (wear with dentine exposure and loss of clinical crown height >1/2-2/3). Twenty-one teeth included in the patient records were scored TWES 4, presenting with dentine exposure and the loss of clinical crown height of >2/3.

Details of the levels of intra-observer agreements (O1) and interobserver agreement (O1-O2) (Kappa scores) for the consecutive scoring of tooth wear applying the TWES on the gypsum cast records and the digital intra-oral scan records that were included in this investigation are presented in Table 3. Table 3 also provides information relating to the comparative evaluation between the use of gypsum cast records and digital greyscale intra-oral scan records with the reproducibility of tooth wear scoring with the TWES. For the grading of the overall occlusal/ incisal surfaces using gypsum cast records, the levels of intra-observer agreements (O1) and interobserver agreements (O1-O2) were significantly higher compared with the agreement in the scoring of the same surfaces using the digital greyscale intra-oral scan records, ($P < .001$) and($P < .001$), respectively. For the grading of the overall buccal and palatal/lingual surfaces, other than significantly higher levels of O1 agreement in the scoring of the buccal surfaces using gypsum cast ($P = .013$) and the O1-O2 agreement in the scoring of the palatal/lingual surfaces with gypsum cast records ($P = .043$), no other significant difference was found between the type of record used on the reliability of scoring with the TWES.

Table 4 provides information about the effect of the type of record on the tooth wear score. This was expressed as the mean difference in the tooth wear grading on gypsum casts and the digital greyscale intra-oral scan records using the TWES. For the overall scores at the occlusal/ incisal surfaces, grading of the gypsum casts culminated in significantly higher TWES scores compared with the use of the digital greyscale intra-oral scan records ($P < .001$; 95% CI = [0.084...0.272]). However, the overall scores at the buccal and palatal/lingual surfaces showed significantly higher values using the digital intra-oral scans records when undertaking tooth wear grading than with the use of gypsum cast records ($P = .009$; 95% CI = [−0.294...0.042] and $P = .001$; 95% CI = [−0.342...0.084]), respectively.

4 | DISCUSSION

This study has reported high levels of agreement (both intra- and interobserver) in the scoring of the occlusal/incisal surfaces using gypsum cast records, applying the 8-point grading scales of the TWES on the dental records of ten randomly selected patients with signs of moderate to severe tooth wear. The superiority of using gypsum cast records compared with digital greyscale intra-oral scan records at the occlusal/incisal surfaces was statistically significant. Moreover, significantly higher tooth wear scores were recorded when applying the gypsum cast records for the grading of the occlusal/ incisal surfaces, whereas the opposite was reported for the buccal/palatal surfaces. As with the present investigation, several previous studies have also reported favourable reliability applying the 8-point

TABLE 3 Intra- and interobserver agreements (Kappa scores) using the TWES on gypsum cast records the and digital intra-oral scans Kappa's Cohen (κ_W) of intra- and interobserver measurements per location (01 = Observer 1, 02 = Observer 2)

		Occlusal/incisal (8-point scale)			Buccal (3-point scale)			Palatal/lingual (3-point scale)		
		Anterior	**Posterior**	**Overall**	**Anterior**	**Posterior**	**Overall**	**Anterior**	**Posterior**	**Overall**
Intra-agreement (O1)	Gypsum	0.86	0.79	0.85	1.00[a]	0.56	0.59	1.00[a]	0.36	0.48
	Digital	0.29	0.54	0.41	0	0.35	0.36	0	0.55	0.62
	P-value	**<.001**	**.002**	**<.001**	N/A	**.040**	**.013**	N/A	.058	.088
	95% CI	[0.454..0.0.698]	[0.091..0.0.403]	[0.343..0.0.541]		[0.010..0.0.412]	[0.050..0.0.420]		[−0.386..0.0.006]	[−0.310..0.0.021]
Interagreement (O1-O2)	Gypsum	0.88	0.59	0.8	1.00[a]	0.13	0.17	0	0.19	0.41
	Digital	0.34	0.22	0.3	0.05	0.12	0.1	0.01	0.24	0.23
	P-value	**<.001**	**<.001**	**<.001**	N/A	.908	.500	.875	.599	**.043**
	95% CI	[0.405..0.0.661]	[0.199..0.0.525]	[0.402..0.0.608]		[−0.191..0.0.214]	[−0.129..0.0.264]	[−0.159..0.0.135]	[−0.253..0.0.146]	[0.006..0.0.348]

Note: Differences between the Kappa scores on gypsum cast records versus digital intra-oral scans, expressed with *P*-value and 95%CI.

N/A: no statistical test was possible.

Bold denotes a value that was statistically significant.

[a]Measurement showed a perfect agreement on a single score.

TABLE 4 Showing the differences in TWES scores between measurements on gypsum casts and digital intra-oral scans

	Occlusal/incisal			Buccal			Palatal/lingual		
	Anterior	Posterior	Overall	Anterior	Posterior	Overall	Anterior	Posterior	Overall
Mean dif.	0.257	0.120	0.178	-0.396	-0.005	-0.168	0.017	-0.387	-0.213
P-value	**.002**	**.037**	**<.001**	**<.001**	.958	**.009**	.836	**<.001**	**.001**
95% CI	[0.096…0.419]	[0.008…0.233]	[0.084…0.272]	[-0.58…-0.212]	[-0.174…0.165]	[-0.294…-0.042]	[-0.141…0.175]	[-0.577…-0.196]	[-0.342…-0.084]

Note: Table is presenting the mean difference between tooth wear gradings on gypsum models versus digital scans using the TWES, together with the *P*-value and the 95%CI.

To test the differences the TWES index was converted into an 8-point-scale (0 = 1, 1a = 2, 1b = 3, 1c = 4, 2 = 5, 3a = 6, 3b = 7, and 4 = 8).

A positive score means a higher tooth wear score on the gysum models compared to the digital scan records.

Bold denotes a value that was statistically significant.

occlusal/incisal grading scale of the TWES for the assessment of worn occlusal/ incisal surfaces using traditional dental casts.[17,19,20] However, with each of these previous investigations, Interclass Correlation coefficients (ICC's) were used to determine reliability. ICC's have been developed for the analysis of continuous outcomes. Furthermore, given that the results of the ICC calculation may be significantly affected with the choice to investigate agreement (as in this case, rather than consistency) the decision was taken to use weighted Kappa scores.

Although tooth wear assessment using the TWES chairside has been shown to be more reliable than assessments carried out using dental casts records alone,[19] reliability investigations have shown the outcomes offered by the use of intra-oral photographs for occlusal/incisal grading to be comparable to the use of gypsum casts.[17] The presence or absence of initial dentine exposure will, however, be more challenging to ascertain using dental casts alone,[9] as the identification of the visual colour changes and subtle tactile alterations at the dental hard tissues that accompany the wear process (and are often associated with the early stages of tooth wear) may not be as readily as detectable compared with chairside assessment. As a limitation of the present study, no patient records were included of cases demonstrating lower levels, or signs of no tooth wear. Furthermore, yellow-coloured Type III dental stone was used for the fabrication of the cast records. Whilst Type III dental stone is intended for the construction of dental casts, the use of a Type IV gypsum material (typically used for the fabrication of dental dies) that can offer higher abrasion resistance and possible finer surface detail, may have had an impact on the observations reported. This may be an area for future investigation, as may be the influence of the colour of the gypsum product on the scoring outcomes.

In the current study, using the gypsum cast records, lower levels of intra- and interobserver agreement were reported with the scoring of tooth wear at the occluding/incisal surfaces of the posterior teeth than at the anterior teeth. Given the practical application of an 8-point ordinal scale for the scoring of the occlusal/ incisal surfaces, with multiple options available and the subtle differences especially between the various sub-scales of the TWES, some variation in the scoring between consecutive assessments (both intra- and interexaminer) is perhaps inevitable.

The results of this study also reported comparatively higher levels intra-observer agreement with the scoring of the posterior teeth compared with the anterior teeth, when applying the digital greyscale intra-oral scan records. This observation was independent of the surface scored. Digital intra-oral scans offer the opportunity for the assessor to view the records in multiple directions and also allow the zooming- in of areas of further interest; however, unlike gypsum casts, they do not permit any tactile assessment. Digital models in greyscale (black-white, as in this investigation), neither permit adequate visualisation of the hard tissue colour changes, which may be relevant for the accurate assessment of less severe patterns of tooth wear, or tooth wear at the non-occluding surfaces of the anterior teeth, as discussed above. Although the use of coloured 3D scans may help improve this aspect and permit the visualisation of exposed dentine, the currently available coloured scans appear to provide a sub-optimal contrast of the tooth surfaces. The need for the visual assessment of the colour changes that accompany the tooth wear process may have accounted for the observations at the anterior teeth buccal and palatal/lingual surfaces included; however, the precise reason of the effect of using digital greyscale intra-oral scans on attaining higher tooth wear score at the anterior buccal and palatal/ lingual surfaces, is not known.

In this investigation, where the buccal and palatal/lingual surfaces were graded using the 3-point ordinal scale of the TWES, in general, lower levels of agreement were described compared with the assessments undertaken at the occlusal/ incisal surfaces. This observation was independent of the type of patient record used. However, some caution needs to be applied with the interpretation of the data attained for the scoring of the buccal and palatal/lingual surfaces, as on occasion, exceptionally high levels of agreement ($\kappa_W = 1.0$) were reported for the anterior teeth included within the sample (Table 3). In general, the Cohen's kappa score requires further consideration if one outcome is extremely dominant and other variables are only encountered sporadically. Furthermore, it may not be appropriate to compare the outcomes in agreement at the differing surfaces using the 8-point ordinal scale at one type of surface and the 3-point scale at another. Previous investigations have also reported considerably lower reliability scores for the grading of non-occlusal/ non-incisal surfaces using the TWES on dental casts.[17,19,20] These findings have been postulated to be accounted by the levels of training the observers may have received to carry out appropriate evaluations at these surfaces, or a possible reflection of a flaw of the TWES grading system itself when applied at such surfaces.[19]

The recoding the TWES into a numerical scale and subsequently analysing the differences between gypsum and digital scores for the purpose of investigating the effect of the type of record on the scoring is an approach that may be questioned. The process of undertaking recoding silently assumes the difference between any two consecutive scales of the TWES to be of the same size. However, this is not necessarily the case. Two alternatives were considered to the approach applied in this investigation. Firstly, an extension of the McNemar test, the McNemar-Bowker test; however, due to the large number of categories in relation to the size of the study, this analysis was not effective. For the second alternative, the Wilcoxon signed-rank test was applied. The latter test, whilst suitable for comparing the gypsum cast and digital intra-oral scan scores, is not able to provide a clinically interpretable estimation of the differences between the scores. However, the Wilcoxon signed-rank test was applied to perform a sensitivity analyses and the outcomes compared with the p-values attained from the paired t tests. In all cases, the *P*-values reported were similar. A situation with one test giving a statistically significant difference and the other test labelling the difference not to be statistically different was not observed. Consequently, the authors considered the more easily interpretable paired *t* test to offer a

level of reliability that was deemed sufficient for the purpose of undertaking analysis.

There are some further limitations with the current study. Previous investigations (often using other clinical tooth wear indices) have reported challenges with the accurate grading of early tooth wear using study casts,[15] clinical photographs[23] or both.[24] The clinical background of the observers has also been shown to influence the outcomes of scoring tooth wear using study casts and photographs.[24] In the present study, both observers were of the same discipline. Furthermore, when considering the effect of the type of the record on the scoring, only the outcomes of a single observer's assessments were used, (O1). The impact of the resolution offered by the intra-oral scanning device used in this investigation is neither known.

Although the merits of occlusal/ incisal grading using the TWES on gypsum casts have been highlighted, compliance with the taking of study models in the primary care sector to monitor wear has been shown to be relatively low.[25] Some caution is also required when undertaking assessments of tooth wear using sequential gypsum casts, due to the risks of distortion of the dental materials used, and the effect of the actual dental material(s) selected.[26] Based on the result of this investigation, it may also be challenging to make accurate comparisons between consecutive gypsum cast records and digital intra-oral scans.

In the future, with increasing popularity for intra-oral scanners in dental practice, some clinicians may preferentially choose to use digital scans/ models for the purpose of the sequential monitoring tooth wear, to overcome the challenges with traditional gypsum models, to include storage. The use of an intra-oral scanner may offer the scope to monitor tooth wear progression consistently and accurately,[27-29] inclusive of the use of subtraction techniques that have been more recently reported.[30] This may help overcome some of the drawbacks commonly associated with the fabrication of gypsum dental casts. However, as there are some clear barriers for the current routine use of intra-oral scanners in the primary care setting (to include, economic factors), the importance of using an appropriate tooth wear index to monitor progression of wear is likely to remain, at least in the short to medium term.

5 | CONCLUSIONS

It was concluded that the scores obtained with the grading scales of the TWES on gypsum casts can offer reliability, especially for the grading of the occlusal/incisal surfaces of teeth with signs of moderate to severe wear. The level of reproducibility offered using digital greyscale intra-oral scan records to carry out tooth wear assessments with the TWES was generally inferior to that offered by the use of gypsum casts.

ACKNOWLEDGMENTS

The authors would like to thank Stephanie Ruiters for her contributions with data collection and measurements undertaken in this study.

CONFLICT OF INTEREST

None of the authors have any conflicts of interest to declare.

AUTHORS' CONTRIBUTION

SB Mehta contributed to data interpretation, drafted and critically revised the manuscript. EM Bronkhorst contributed to design and data interpretation, performed all statistical analyses and critically revised the manuscript. L. Crins contributed to data interpretation and critically revised the manuscript. P. Wetselaar contributed to conception, design and data collection, and critically revised the manuscript. MCDNJM Huysmans contributed to conception, design and data interpretation, and critically revised the manuscript. BAC Loomans is the project leader of the Radboud Tooth Wear Project, contributed to conception, design, enrolment of patients, data acquisition and interpretation, and critically revised the manuscript. All authors gave their final approval and agree to be accountable for all aspects of the work.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available from the corresponding author, SBM upon reasonable request.

ORCID

Shamir B. Mehta https://orcid.org/0000-0002-9674-2492
Peter Wetselaar https://orcid.org/0000-0002-9443-1260

REFERENCES

1. Schlueter N, Luka B. Erosive tooth wear - a review on global prevalence and on its prevalence in risk groups. *Br Dent J*. 2018;224(5):364-370. https://doi.org/10.1038/sj.bdj.2018.167
2. Wazani BE, Dodd MN, Milosevic A. The signs and symptoms of tooth wear in a referred group of patients. *Br Dent J*. 2012;213(6):E10. https://doi.org/10.1038/sj.bdj.2012.840
3. Loomans B, Opdam N, Attin T, et al. Severe tooth wear: European consensus statement on management guidelines. *J Adhes Dent*. 2017;19(2):111-119. https://doi.org/10.3290/j.jad.a38102
4. Papagianni CE, van der Meulen MJ, Naeije M, Lobbezoo F. Oral health-related quality of life in patients with tooth wear. *J Oral Rehabil*. 2013;40(3):185-190. https://doi.org/10.1111/joor.12025
5. Li MHM, Bernabé E. Tooth wear and quality of life among adults in the United Kingdom. *J Dent*. 2016;55:48-53. https://doi.org/10.1016/j.jdent.2016.09.013
6. Sterenborg BAMM, Bronkhorst EM, Wetselaar P, Lobbezoo F, Loomans BAC, Huysmans MDNJM. The influence of management of tooth wear on oral health-related quality of life. *Clin Oral Investig*. 2018;22(7):2567-2573. https://doi.org/10.1007/s00784-018-2355-8
7. O'Toole S, Pennington M, Varma S, Bartlett DW. The treatment need and associated cost of erosive tooth wear rehabilitation - a service evaluation within an NHS dental hospital. *Br Dent J*. 2018;224(12):957-961. https://doi.org/10.1038/sj.bdj.2018.444
8. Mehta SB, Banerji S, Millar BJ, Suarez-Feito JM. Current concepts on the management of tooth wear: part 1. Assessment, treatment planning and strategies for the prevention and the passive management of tooth wear. *Br Dent J*. 2012;212(1):17-27. Published 2012 Jan 13. doi:10.1038/sj.bdj.2011.1099.
9. Bartlett DW. Retrospective long term monitoring of tooth wear using study models. *Br Dent J*. 2003;194(4):211-213. https://doi.org/10.1038/sj.bdj.4809914

10. Bardsley PF. The evolution of tooth wear indices. *Clin Oral Investig*. 2008;12(Suppl 1):S15-S19. https://doi.org/10.1007/s00784-007-0184-2
11. Smith BG, Knight JK. An index for measuring the wear of teeth. *Br Dent J*. 1984;156(12):435-438. https://doi.org/10.1038/sj.bdj.4805394
12. Bartlett D, Ganss C, Lussi A. Basic Erosive Wear Examination (BEWE): a new scoring system for scientific and clinical needs. *Clin Oral Investig*. 2008;12(Suppl 1):S65-S68. https://doi.org/10.1007/s00784-007-0181-5
13. Lussi A, Schaffner M, Hotz P, Suter P. Dental erosion in a population of Swiss adults. *Community Dent Oral Epidemiol*. 1991;19(5):286-290. https://doi.org/10.1111/j.1600-0528.1991.tb00169.x
14. Wetselaar P, Lobbezoo F. The tooth wear evaluation system (TWES): a modular clinical guideline for the diagnosis and management planning of worn dentitions. *J Oral Rehabil*. 2016;43:69-80.
15. Wetselaar P, Faris A, Lobbezoo F. A plea for the development of an universally accepted modular tooth wear evaluation system. *BMC Oral Health*. 2016;16(1):115.
16. Dixon B, Sharif MO, Ahmed F, Smith AB, Seymour D, Brunton PA. Evaluation of the basic erosive wear examination (BEWE) for use in general dental practice. *Br Dent J*. 2012;213(3):E4. https://doi.org/10.1038/sj.bdj.2012.670
17. Wetselaar P, Wetselaar-Glas MJ, Koutris M, Visscher CM, Lobbezoo F. Assessment of the amount of tooth wear on dental casts and intra-oral photographs. *J Oral Rehabil*. 2016;43(8):615-620. https://doi.org/10.1111/joor.12405
18. Wetselaar P, Wetselaar-Glas MJM, Katzer LD, Ahlers MO. Diagnosing tooth wear, a new taxonomy based on the revised version of the tooth wear evaluation system (TWES 2.0). *J Oral Rehabil*. 2020;47(6):703-712. https://doi.org/10.1111/joor.12972
19. 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(4):388-390.
20. Vervoorn-Vis GM, Wetselaar P, Koutris M, et al. Assessment of the progression of tooth wear on dental casts. *J Oral Rehabil*. 2015;42(8):600-604. https://doi.org/10.1111/joor.12292
21. Loomans B, Opdam N. A guide to managing tooth wear: the Radboud philosophy. *Br Dent J*. 2018;224(5):348-356. https://doi.org/10.1038/sj.bdj.2018.164
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
23. Mulic A, Tveit AB, Wang NJ, Hove LH, Espelid I, Skaare AB. Reliability of two clinical scoring systems for dental erosive wear. *Caries Res*. 2010;44(3):294-299. https://doi.org/10.1159/000314811
24. Paesani DA, Guarda-Nardini L, Gelos C, Salmaso L, Manfredini D. Reliability of multiple-degree incisal/occlusal tooth wear assessment on dental casts: findings from a fiveexaminer investigation and related clinical implications. *Quintessence Int*. 2014;45(3):259-264. https://doi.org/10.3290/j.qi.a31212
25. Bartlett DW, Palmer I, Shah P. An audit of study casts used to monitor tooth wear in general practice. *Br Dent J*. 2005;199(3):143-145. https://doi.org/10.1038/sj.bdj.4812570
26. Tarawneh FM, Panos PG, Athanasiou AE. Three-dimensional assessment of dental casts' occlusal surfaces using two impression materials. *J Oral Rehabil*. 2008;35(11):821-826. https://doi.org/10.1111/j.1365-2842.2008.01872.x
27. Ahmed KE, Whitters J, Ju X, Pierce SG, MacLeod CN, Murray CA. A proposed methodology to assess the accuracy of 3D scanners and casts and monitor tooth wear progression in patients. *Int J Prosthodont*. 2016;29(5):514-521. https://doi.org/10.11607/ijp.4685
28. Ahmed KE, Whitters J, Ju X, Pierce SG, MacLeod CN, Murray CA. Clinical monitoring of tooth wear progression in patients over a period of one year using CAD/CAM. *Int J Prosthodont*. 2017;30(2):153-155. https://doi.org/10.11607/ijp.4990
29. Kumar S, Keeling A, Osnes C, Bartlett D, O'Toole S. The sensitivity of digital intraoral scanners at measuring early erosive wear. *J Dent*. 2019;81:39-42. https://doi.org/10.1016/j.jdent.2018.12.005
30. O'Toole S, Lau JS, Rees M, Warburton F, Loomans B, Bartlett D. Quantitative tooth wear analysis of index teeth compared to complete dentition. *J Dent*. 2020;97:103342. https://doi.org/10.1016/j.jdent.2020.103342

How to cite this article: Mehta SB, Bronkhorst EM, Crins L, Huysmans M-CDNJ, Wetselaar P, Loomans BAC. A comparative evaluation between the reliability of gypsum casts and digital greyscale intra-oral scans for the scoring of tooth wear using the Tooth Wear Evaluation System (TWES). *J Oral Rehabil*. 2021;48:678–686. https://doi.org/10.1111/joor.13141