Reliability of Two Clinical Scoring Systems for Dental Erosive Wear

A. Mulic\textsuperscript{a} A.B. Tveit\textsuperscript{a} N.J. Wang\textsuperscript{b} L.H. Hove\textsuperscript{a} I. Espelid\textsuperscript{b} A.B. Skaare\textsuperscript{b}

\textsuperscript{a}Department of Cariology and \textsuperscript{b}Department of Pediatric Dentistry and Behavioural Science, Faculty of Dentistry, University of Oslo, Oslo, Norway

Key Words
Dental erosion · Reliability · Scoring systems

Abstract
The aim of the study was to evaluate and compare two dental erosive wear scoring systems, the Visual Erosion Dental Examination (VEDE) and Basic Erosive Wear Examination (BEWE). Seventy-four tooth surfaces (photographs) and 562 surfaces (in participants) were scored by 5 (photographs) or 3 (in participants) clinicians using both scoring systems. The surfaces in the photographs were scored twice. The level of agreement was measured by weighted kappa ($\kappa_w$). Inter- and intraexaminer agreement showed small variations between the examiners for both systems when scoring the photographs. Slightly higher mean $\kappa_w$ values were found for VEDE ($\kappa_w = 0.77$) compared with BEWE ($\kappa_w = 0.69$). When scoring the surfaces in the clinical examination the mean $\kappa_w$ values for the two systems were equal ($\kappa_w = 0.73$). Interexaminer agreement using VEDE was calculated to see how differentiation between enamel and dentine lesions influenced the variability. The highest agreement was found for score 0 (sound, 86%) and score 3 (exposure of dentine, 67%), while the smallest agreement was shown for score 1 (initial loss of enamel, 30%) and score 2 (pronounced loss of enamel, 57%). The reliability of the two scoring systems proved acceptable for scoring the severity of dental erosive wear and for recording such lesions in prevalence studies. The greatest difficulties were found when scoring enamel lesions, especially initial lesions, while good agreement was observed when examining sound surfaces (score 0) and dentine lesions (score 3).

During the last two decades, dental erosive wear has attracted considerable attention. In many epidemiological studies, high prevalence of such wear has been found [Milosevic et al., 1994; al-Dlaigan et al., 2001; Al-Majed et al., 2002; Dugmore and Rock, 2004; El Aidi et al., 2008]. However, wide variation was reported. An Icelandic survey [Arnadottir et al., 2003] found that 21.6% of 15-year-olds suffered from dental erosions, while Al-Majed et al. [2002] reported a prevalence of 95% in the permanent dentition among 12- to 14-year-old Saudi Arabian boys. Differences could be explained by the fact that the age groups included varied: some studies report tooth erosion in primary and mixed dentitions [Al-Majed et al., 2002; Wiegand et al., 2006], while others report erosion in permanent teeth in adolescents [Milosevic et al., 1994; Al-Majed et al., 2002; Arnadottir et al., 2003; Dugmore and Rock, 2004] or in adults [Lussi et al., 1991; Johansson et al., 1996; Fares et al., 2009]. The variation in prevalence may also be due to different dietary habits and lifestyle in different countries. For instance, in Saudi Arabia, with its extremely hot climatic conditions, the consumption of soft drinks ex-
ceeds the consumption normally found in Western populations [Johansson et al., 1997]. Furthermore, while some authors assess dental erosion on all surfaces [Lussi et al., 1991; Milosevic et al., 1994; Larsen et al., 2000, 2005; El Aidi et al., 2008], others grade hard tissue loss on first molars and incisors [Nunn et al., 2003; Bardsley et al., 2004; Dugmore and Rock, 2004], or on incisors only [Johansson et al., 1996; Williams et al., 1999].

Another major factor which must be taken into account when comparing data from different epidemiological studies is the different scoring systems used to assess tooth substance loss [Bardsley, 2008]. Unlike the study by Larsen et al. [2000], most epidemiological studies have been carried out with measurement systems of untested reliability. The reliability of the measuring instrument should be known or, if necessary, evaluated before initiating a study of dental erosive wear. The Visual Erosion Dental Examination (VEDE) system has been used in the student clinics at the University of Oslo for 5 years (fig. 1, tables 1, 2), and came to agreement. The remaining 74 photographs were used in the comparison of the two systems.

Two separate examination sessions were arranged for scoring the severity of dental erosive wear. VEDE was used in the first session and BEWE in the second, 1 month later. For both scoring systems the assessments were repeated after 14 days. When there was discrepancy between the scores given on these separate occasions, the examiners chose the lower score. All examinations were conducted in the same room and under identical lighting conditions.

Clinical Examination
Three of the 5 clinicians examined thirty 18-year-old adolescents who had been referred to the Dental School at the University of Oslo, from the Public Dental Services because of clear or suspected signs of dental erosive wear.

Written, informed consent was obtained from all participants. The study was approved by the local Regional Committee for Medical Research Ethics and The Norwegian Social Science Data Services.

Table 1. VEDE system

| Score | Definition |
|-------|------------|
| 0     | No erosion |
| 1     | Initial loss of enamel, no dentine exposed |
| 2     | Pronounced loss of enamel, no dentine exposed |
| 3     | Exposure of dentine, <1/3 of the surface involved |
| 4     | 1/3–2/3 of the dentine exposed |
| 5     | >2/3 of dentine exposed, or pulp exposed |

Table 2. BEWE system

| Score | Definition |
|-------|------------|
| 0     | No erosion |
| 1     | Initial loss of surface texture |
| 2a    | Distinct defect, hard tissue loss less than 50% of the surface area |
| 3a    | Hard tissue loss more than 50% of the surface area |

a Dentine is often involved.

Material and Methods

Photographic Examination
A total of 104 intraoral close-up photographs of tooth surfaces from patients were randomly selected from the University Dental School’s patient photo archive to depict different groups of teeth, surfaces and severity grades of erosive tooth wear. Each photograph showed one tooth surface. About one fifth of these surfaces were considered to be sound and the surfaces with dental erosive wear covered the whole range from minor to obviously severe lesions. Five examiners with dental professional experience from 3 to 35 years and working with children, adolescents and adults examined the pictures. Based on a subsample of 30 photographs, the examiners discussed the criteria of VEDE and BEWE (fig. 1, tables 1, 2), and came to agreement. The remaining 74 photographs were used in the comparison of the two systems.

Before initiating a study of dental erosive wear it would be desirable to assess the reliability of the measuring systems to be used. The aim of the present study was therefore to evaluate and compare two established clinical measuring systems for erosive dental wear, VEDE and BEWE.
Examination took place in a dental clinic with standard lighting and using mouth mirrors and probes. Surfaces were dried by compressed air and, if necessary, cotton rolls were used to remove food debris prior to the examinations. Each examination included 20 surfaces per participant: the occlusal surfaces of the first and second permanent molars in both jaws and on the labial and palatal surfaces of the upper incisors and canines. Examiners made individual recordings. Participants were examined twice during an examination, first with the VEDE and, approximately 15 min later, with the BEWE. Only lesions that were considered to be obvious dental erosive wear defects were recorded and scored, including 'pits'/grooves of the molar cusps. Attritions and wedge-shaped defects on occlusal and incisal surfaces were not graded.

Fig. 1. VEDE system.
To calculate intraexaminer agreement, 15 adolescents were re-examined by one of the examiners 10–21 days after the first examination.

**Statistical Analysis**

The reliability of the scoring systems was assessed measuring inter- and intraexaminer agreement by linear weighted Cohen's kappa ($\kappa_w$). The statistical analysis for the weighted kappa was calculated using a spreadsheet program (Microsoft Excel). Cohen's kappa was rated as suggested by Landis and Koch [1977]: $<0.40$, $0.41–0.60$, $0.61–0.80$, $0.81–1.0$.

When index surfaces were filled, bonded with a retainer, considered to have attritions and wedge-shaped defects or the tooth was extracted, the surfaces and teeth were recorded as missing ($n = 38$) and excluded.

**Results**

The distribution of the scores recorded in the photographic and clinical examination is presented in tables 3 and 4. The total number of surfaces on chosen index surfaces was 74 and 562 for the photographic and clinical session, respectively. No participants presented with VEDE dental erosive wear score 5, but all other scores were used in the clinical session. Most surfaces were registered as score 0, 1 and 2 (tables 3, 4).

Interexaminer agreement varied more when using BEWE than VEDE when the photographs were assessed. Mean $\kappa_w$ values were higher for VEDE than for BEWE. The values of $\kappa_w$ for both systems indicated acceptable agreement. The same pattern was observed for the intraexaminer measurement, but the $\kappa_w$ values were higher (table 5).

$\kappa_w$ values were equal for both scoring systems in the clinical examination and indicated substantial agreement. Variation between examiners in the clinical session was similar to that seen when examining photographs. Intraexaminer agreement ($\kappa_w$), when 15 patients were re-examined, was 0.92 and 0.95 for BEWE and VEDE, respectively.

To find out whether interexaminer variation differed when scoring enamel lesions compared with lesions in dentine using VEDE, percentage agreement (%) was cal-

---

Table 3. Photographic examination

| Erosion score: | 0 | 1 | 2 | 3 | 4 | 5 | Total |
|---------------|---|---|---|---|---|---|-------|
| Scoring system: | VEDE | BEWE | VEDE | BEWE | VEDE | BEWE | VEDE | BEWE | VEDE | VEDE |
| Examiner A | 15 | 12 | 7 | 9 | 17 | 10 | 13 | 33 | 10 | 12 | 74 |
| B | 12 | 13 | 10 | 10 | 17 | 16 | 9 | 35 | 14 | 12 | 74 |
| C | 9 | 11 | 4 | 8 | 21 | 13 | 21 | 42 | 10 | 9 | 74 |
| D | 13 | 15 | 5 | 7 | 18 | 18 | 10 | 34 | 13 | 15 | 74 |
| E | 14 | 16 | 7 | 8 | 18 | 11 | 10 | 39 | 12 | 13 | 74 |
| All examiners | 63 | 67 | 33 | 42 | 91 | 78 | 68 | 183 | 59 | 61 | 370 |

Distribution of dental erosive scores on tooth surfaces using VEDE and BEWE systems by 5 examiners (A–E) ($n = 74$).

Table 4. Clinical examination

| Erosion score: | 0 | 1 | 2 | 3 | 4 | 5 | Total |
|---------------|---|---|---|---|---|---|-------|
| Scoring system: | VEDE | BEWE | VEDE | BEWE | VEDE | BEWE | VEDE | BEWE | VEDE | VEDE |
| Examiner C | 421 | 428 | 45 | 40 | 80 | 60 | 11 | 34 | 5 | 562 |
| D | 428 | 431 | 49 | 47 | 67 | 50 | 14 | 34 | 3 | 562 |
| E | 453 | 451 | 17 | 20 | 73 | 47 | 15 | 44 | 4 | 562 |
| All examiners | 1,303 | 1,310 | 111 | 107 | 220 | 157 | 40 | 112 | 12 | 1,686 |

Distribution of dental erosive scores on tooth surfaces using VEDE and BEWE systems by 3 examiners (C–E) ($n = 562$).
calculated. In the clinical situation the examiners agreed on 86% of the surfaces scored 0, and on 30, 57 and 67% of the surfaces scored 1, 2 and 3, respectively.

**Discussion**

In the present study, examiner reliability of two different scoring systems, expressed by $k_w$, was shown to be acceptable when scoring the severity of erosive tooth wear. In a review by Bardsley [2008], a variety of tooth wear indices was evaluated, and the author pointed out some restrictions associated with the systems. According to this review, the lack of standardization of terminology and vague definitions of criteria are limitations which allow a wide interpretation of severity scores. Larsen et al. [2005] concluded in their study that the classification systems applied for erosion diagnosis are difficult to use, mainly because dental erosion does not change enamel color, which can complicate the identification of the scores. Dental practitioners and investigators in this research field are familiar with these limitations.

Since both scoring systems apply score 1 to define the earliest or the initial erosive enamel lesions, the question arises whether photographs are a suitable tool for detecting these minor lesions. There are, as far as we know, no studies reporting the reliability of scoring on photographs compared with clinical examination. The present results show a reasonable spread of data between scores 0 and 1 (table 3). This indicates that the use of picture-based classification could be acceptable for measuring dental erosive wear and may be on a par with clinical examination.

The BEWE system does not distinguish between enamel loss and exposed dentine, which could be regarded as a way to avoid diagnostic uncertainties [Bartlett et al., 2008]. This view is shared by Ganss et al. [2006], who concluded that the visual diagnosis of exposed dentine is difficult. VEDE has two scores for enamel loss and a detailed distinction between dentine scores (scores 3, 4 and 5), which did not seem to influence the variability between observers compared with BEWE. A higher proportion of agreement was found between the examiners for VEDE dentine score 3 (67%), compared with enamel scores 1 (30%) and 2 (57%). This system may not support diagnostic uncertainties regarding dentine diagnosis, as pointed out by Ganss et al. [2006] and Bartlett et al. [2008]. One should bear in mind that the differentiation between enamel and dentine may be an important factor for recording progression of dental erosion and wear [Ganss et al., 2006; Fares et al., 2009] and in treatment planning. Another advantage in defining dentine exposure is its wide use by most scoring systems, making the results from other studies easier to compare.

In the BEWE system, scores 2 and 3 are broadly defined as they cover tissue loss from a minimal, but distinct, loss of surface (less than 50%) to the loss of almost all enamel and/or dentine (more than 50%). For the VEDE system, the weakness may be its detailed scale, especially the distinction between no dental erosive wear, initial loss of enamel (score 1) and pronounced loss of enamel (score 2). This could influence the reproducibility of the scores as shown by interexaminer agreement levels in the clinical examination, which was only 30% for score 1. However, this study does not show whether the low examiner agreement is due to the system’s detailed scoring scale or the examiner’s weakness in recording the initial lesions in general. In the study by Larsen et al. [2005], the examiners also had difficulties handling the limit between intact enamel and early enamel lesion.

The clinical examination was performed on index teeth and surfaces. The selection of the surfaces was based on earlier ‘full mouth recording’ studies among adolescents [Milosevic et al., 1994; al-Dlaigan et al., 2001; Ganss et al., 2001; van Rijkom et al., 2002; Larsen et al., 2005], which demonstrated the highest prevalence of dental erosion on occlusal surfaces of molars and labial and palatal surfaces of maxillary anterior teeth. Since scoring was performed only on index teeth and surfaces, the present study deviated from BEWE’s original requirement to examine all teeth in a quadrant.

In the article by Larsen et al. [2005], where 9 clinicians examined the same subjects, low interexaminer agree-
ment was observed. The average number of eroded surfaces ranged from 4.1 to 21.3. However, the results from the present study, demonstrated by the \( \kappa \) values, showed acceptable variation between the 5 examiners (table 5). One explanation for the difference in the level of agreement between the two studies may be found in the calibration of examiners and the definition of the categories of the scoring systems. On the other hand, the lower number of examiners contributing to the present study could have reduced variability compared with the study by Larsen et al. [2005].

It is difficult to compare the reliability of different studies due to the variety of scoring systems used. The results in the present study indicate that a scoring system should clearly define categories of dental erosive wear and that the use of pictures may assist in the scoring. A system that differentiates between erosive tooth wear in enamel and dentine is valuable for prognostic purposes and does not seem to reduce the variability between observers. These remarks lead to the conclusion that the two systems we tested may have different applications. While the BEWE system seems suitable for clinical screening for epidemiological purposes because it has fewer categories, the VEDE system’s strength lies in its ability to diagnose the early stages of the condition and to record progression of tooth erosive wear on an individual basis. Validating the scoring systems prior to an epidemiological study seems important and may be an attempt to avoid diagnostic uncertainties. It should also be noted that early diagnosis of dental erosive wear and its progression are important for counseling and informative purposes, and will make treatment planning easier for the patient and the clinician. There is a need in future research to explore methods for validation of erosive wear systems, to record correctly loss of tooth substance.

References

- Al-Dlaigan YH, Shaw L, Smith A: Dental erosion in a group of British 14-year-old, school children. Part I: Prevalence and influence of differing socioeconomic backgrounds. Br Dent J 2001;190:145–149.
- Al-Majed I, Maguire A, Murray JJ: Risk factors for dental erosion in 5–6 year old and 12–14 year old boys in Saudi Arabia. Community Dent Oral Epidemiol 2002;30:38–46.
- Arnadottir IB, Saemundsson SR, Holbrook WP: Dental erosion in Icelandic teenagers in relation to dietary and lifestyle factors. Acta Odontol Scand 2003;61:25–28.
- Bardsey PF: The evolution of tooth wear indices. Clin Oral Investig 2008;12(suppl 1):S15–S19.
- Bardsey PF, Taylor S, Milosevic A: Epidemiological studies of tooth wear and dental erosion in 14-year-old children in North West England. Part I: The relationship with water fluoridation and social deprivation. Br Dent J 2004;197:413–416.
- 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.
- Bugis Я, Rock WP: The prevalence of tooth erosion in 12-year-old children. Br Dent J 2004;196:279–282.
- El Aidi H, Bronkhorst EM, Truin GJ: A longitudinal study of tooth erosion in adolescents. J Dent Res 2008;87:731–735.
- Fares J, Shirodaria S, Chiu K, Ahmad N, Sherriff M, Bartlett D: A new index of tooth wear: reproducibility and application to a sample of 18– to 30-year-old university students. Caries Res 2009;43:119–125.
- Ganss C, Klimek J, Giese K: Dental erosion in children and adolescents: a cross-sectional and longitudinal investigation using study models. Community Dent Oral Epidemiol 2001;29:264–271.
- Ganss C, Klimek J, Lussi A: Accuracy and consistency of the visual diagnosis of exposed dentine on worn occlusal/incisal surfaces. Caries Res 2006;40:208–212.
- Johannson AK, Johannsson A, Birkhed D, Omar R, Baghdadi S, Carlson GE: Dental erosion, soft-drink intake, and oral health in young Saudi men, and the development of a system for assessing erosive anterior tooth wear. Acta Odontol Scand 1996;54:369–378.
- Johannson AK, Johannsson A, Birkhed D, Omar R, Baghdadi S, Khan N, Carlson GE: Dental erosion associated with soft-drink consumption in young Saudi men. Acta Odontol Scand 1997;55:390–397.
- Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
- Larsen JB, Westergaard J, Stoltze K, Larsen AI, Gyntelberg F, Holmstrup P: A clinical index for evaluating and monitoring dental erosion. Community Dent Oral Epidemiol 2000;28:211–217.
- Larsen MJ, Poulsen S, Hansen I: Erosion of the teeth: prevalence and distribution in a group of Danish school children. Eur J Paediatr Dent 2005;6:44–47.
- Lussi A: Dental erosion clinical diagnosis and case history taking. Eur J Oral Sci 1996;104:191–198.
- Lussi A, Schaffner M, Hotz P, Suter P: Dental erosion in a population of Swiss adults. Community Dent Oral Epidemiol 1991;19:286–290.
- Milosevic A, Young PJ, Lennon MA: The prevalence of tooth wear in 14-year-old school children in Liverpool. Community Dent Health 1994;11:83–86.
- Nunn JH, Gordon PH, Morris AJ, Pine CM, Walker A: Dental erosion – changing prevalence? A review of British National children’s surveys. Int J Paediatr Dent 2003;13:98–105.
- van Rijkom HM, Truin GJ, Frencken JE, König KG, van ‘t Hof MA, Bronkhorst EM, Roeters FJ: Prevalence, distribution and background variables of smooth-bordered tooth wear in teenagers in the Hague, the Netherlands. Caries Res 2002;36:147–154.
- Wiegand A, Muller J, Werner C, Attin T: Prevalence of erosive tooth wear and associated risk factors in 2–7-year-old German kindergarten children. Oral Dis 2006;12:117–124.
- Williams D, Croucher R, Marcenes W, O’Farrell M: The prevalence of dental erosion in the maxillary incisors of 14-year-old schoolchildren living in Tower Hamlets and Hackney, London, UK. Int Dent J 1999;49:211–216.
- Young A, Amaechi BT, Bugis Я, Holbrook P, Nunn J, Schiffner U, Lussi A, Ganss C: Current erosion indices: flawed or valid? Summary. Clin Oral Investig 2008;12(suppl 1):S59–S63.