

IN BRIEF

- QLF is a new device to detect very early demineralisation of enamel.
- QLF can be used *in vivo* and *in vitro*.
- All diagnostic methods suffer from some form of variability.
- This study determined that, in most cases, QLF does not suffer from significant reliability issues in the analysis stage.
- Further work on *in vivo* lesions is required.
- Novices should be trained before analysing experimental data.

The intra- and inter-examiner reliability of quantitative light-induced fluorescence (QLF) analyses

I. A. Pretty,¹ A. F. Hall,² P. W. Smith,³ W. M. Edgar⁴ and S. M. Higham⁵

Objective To assess the reliability of the analysis stage of quantitative light-induced fluorescence (QLF). The QLF analysis involves subjective input from the user and this study examines the influence of this on the reproducibility of the QLF data.

Method QLF images were taken of 20 human molar teeth that had been previously subjected to a demineralising solution (phosphoric acid 37%) to create artificial white spot lesions on their buccal surfaces. Following examination of the images, 16 were chosen to represent a range of lesion size and severity. Three copies were made of the images and each was allocated a different filename. 10 examiners in three centres were asked to analyse each of the 16 images on three occasions, with at least seven days between each attempt. Simple instructions describing the analysis procedure were supplied and examiners were asked to adhere to these directions. Examiners were asked to rate each of the 16 teeth on their first attempt both quantitatively (5 point scale) and qualitatively in terms of difficulty of analysis. Data reported were the ΔQ at 5% threshold for each tooth on each of three attempts.

Results Using ANOVA and paired *t*-tests to detect statistical differences, the three attempts of each examiner were used to determine intra-examiner reliability. Only one examiner (a novice at the technique) demonstrated differences between all three attempts and two demonstrated difference between one attempt. When the mean scores were compared to determine the inter-examiner reliability, only one examiner's results were statistically different when compared with two others.

Conclusion This study has demonstrated that the analysis stage of QLF is reliable between examiners and within multiple attempts by the same examiner, when analysing *in vitro* lesions. Novices at the technique should be trained before analysing experimental data.

Quantitative light-induced fluorescence (QLF) is a novel technique for the detection of early carious lesions in human enamel. Based upon the detection, and quantifiable analysis of an increased scattering of fluorescent light associated with demineralised lesions, the QLF process involves two main stages. The first stage is the capture of the tooth image. Following this the resultant image must be analysed using the proprietary software supplied with the device to produce the final result, a measure of mineral loss. The purpose of this study is to investigate the reliability, based upon intra- and inter-examiner agreement, of this stage of QLF analysis.

The QLF device (Inspektor Research Systems, bv, The Netherlands) can be considered as a new measurement device, and, as such, it must be subjected to a variety of tests to ensure that it is both valid and reliable. There are a number of methods of achieving this. First, validity, which can best be described as 'is this device measuring what it claims to measure?' The most usual method of testing this is to compare the new method with the current gold standard within the field. This has been studied by cross-validating QLF with Transverse Micro-Radiography (TMR) that found strong correlations between the two techniques.¹⁻³

When assessing the reliability of the instrument, one must consider if there is likely to be any operator bias. The QLF device has the potential for operator bias, as it relies upon a subjective analysis of the stored tooth image. An example of another optical device, the KaVo DIAGNODent,⁴ does not suffer from such a potential bias, as a value is displayed on a LCD panel without requiring further input from the operator. However, the increased data yield and quantifiable nature of the QLF value, ΔQ , dictates that a more complex analysis is required. The QLF device, rather than merely supplying the arbitrary value characteristic of the DIAGNODent, supplies information on fluorescence loss (ΔF), lesion size (mm^2) and the integration of these two values, ΔQ . The ΔQ value is therefore an indicator of mineral loss with relation to size and severity of the lesion.

The method of QLF analysis is described in detail within the methods and materials section, however the subjectivity of the technique is discussed here. The QLF method relies upon the subjective determination by the operator of several aspects of the QLF image:

¹Graduate student, The University of Liverpool, Department of Clinical Dental Sciences;

²Senior Lecturer, The University of Glasgow, Department of Restorative Dentistry;

³Lecturer, The University of Manchester, Department of Prosthodontics; ⁴Emeritus Professor, Senior Research Fellow, The University of Liverpool, Clinical Dental Sciences;

⁵Senior Lecturer, The University of Liverpool, Department of Clinical Dental Sciences

*Correspondence to: I. A. Pretty, The University of Liverpool, Department of Clinical Dental Sciences, Edwards Building, Daulby Street, Liverpool L69 3GN
E-mail: ipretty@liv.ac.uk

Refereed paper

Received 07.06.01; Accepted 22.02.02

© British Dental Journal 2002; 193: 105–109

- Determination of sound or unsound enamel.
- Determination of the presence or absence of a demineralised lesion.
- The borders of any detected lesion.
- The morphology of the lesion.
- Presence of any confounding factors, ie plaque, calculus, moisture, shadowing, enamel defects etc.

Because of these subjective components of QLF an investigation into the reliability of the technique is required to gauge their influence upon the validity of the measurement.

Measurement variability

With any measurement there will be variability. For a measurement of any single factor, repeated measurements of that same factor will vary. Fleiss⁵ describes this; a true value (T) could be considered as one that is a mean of many repeated measurements. Subsequent individual measurements (X) will vary from this true value for any number of different reasons, including

examiner effects, subject effects, imperfect calibration, etc. The true value (T) will vary around a given mean, and the random error (E) of the individual measurement (X) will vary around a mean of zero. This is shown in the following formula:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

From this it is possible to determine that a variation of the ΔQ value following repeated measures of the same lesion is inevitable. The purpose of this study is to quantify the intra- and inter-examiner reliability of QLF analyses. Such data is essential to researchers using the device and those taking note of findings from such studies. It is necessary to determine that any differences detected by examiners are the result of a manipulated experimental variable rather than statistical artefacts of an unreliable device. It is also important to determine whether different researchers' results can be directly compared. Reliability of a device also impacts on the number of subjects that are required for a trial, and in clinical situations, the weight that is given to a positive or negative diagnostic finding.

METHODS AND MATERIALS

The QLF method

The basic method for analysing a QLF image is described and shown in Figure 1. The technique is open to a great deal of variation in terms of where the patch is placed, the subjective judgement of the operator as to what is sound enamel and which, if any, borders should be excluded. The inclusion or exclusion of borders can have a profound impact on the reported value. However, it is important to note that there are several guides to successful analysis within the program. One of these is the reconstructed image. Following analysis of the image, the reconstructed image is shown adjacent to the original image. The re-constructed image should be clear of any streaking or dark lines, indicating that the lesion has been successfully 'removed' from the reconstruction. Figure 2 shows both a successful and an unsuccessful reconstruction of the same lesion. Another means by which operators can confirm their analysis technique is the reconstructed lesion that is also shown following analysis. Adjacent to the grayscale reconstructed image, an image of the outline of the lesion is shown. This outline should be compared with the original lesion shown on the QLF image. If the analysis has been successful the isolated lesion will closely match the morphology of that present on the original captured image. An example of this is shown in Figure 3.

Lesion and image production

Twenty previously extracted human third molar teeth were cleaned of all soft-tissue remnants, and their buccal surfaces gently pumiced and abraded with wet-and-dry paper. Each tooth was given two coats of acid-resistant nail varnish (MaxFactor, Proctor and Gamble, UK) leaving an exposed window of enamel on the buccal surface. The exposed windows were of varying sizes and position. The teeth were subjected to varying exposures of 37% phosphoric acid gel (Henry Schein, UK) ranging from 5 seconds to 10 minutes. This produced simulated white-spot lesions of varying severity. Etched enamel lesions appear exactly as other simulated lesions under QLF conditions, for example those created by acidic buffers or gels. Each tooth was examined using QLF. Following this examination, 16 teeth were selected to provide a spread of lesion severity, size, position and hence analysis difficulty (Fig. 4). In studies examining examiner reliability it is desirable to achieve an even spread of easy, moderate and complex cases to ensure a valid result. This was initially assessed subjectively by the investigators and then confirmed by asking the examiners to rate each lesion's analysis complexity.

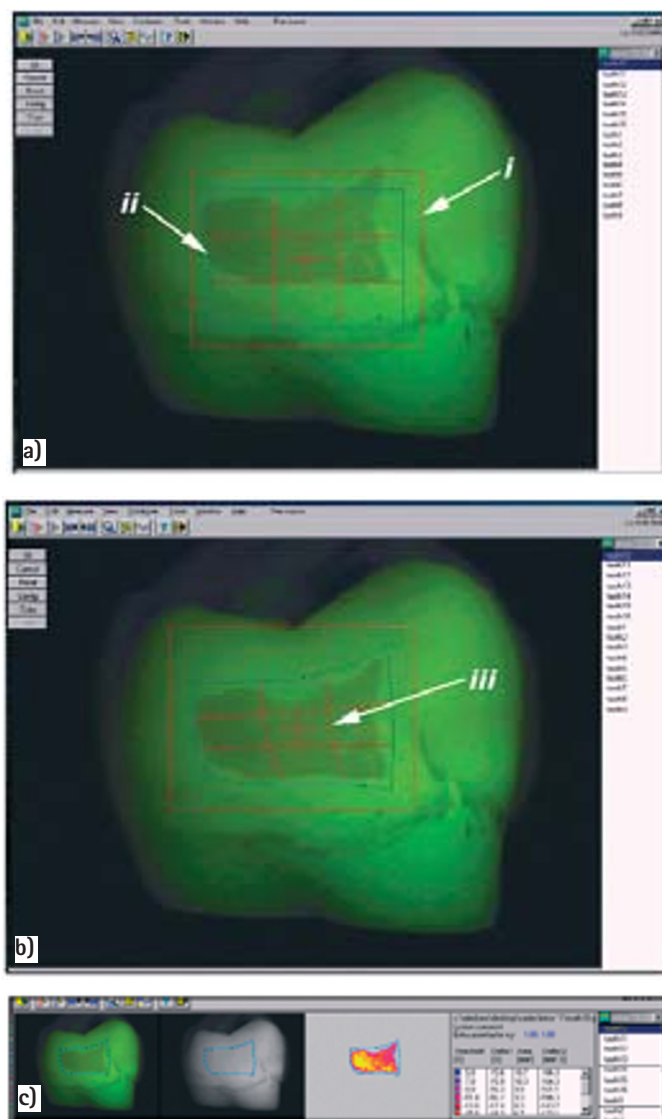


Fig. 1 An example of a QLF analysis. The captured image is shown in a). The solid square (i) is used to control the overall size of the analysis patch (the dotted line (ii)); b) The analysis patch is placed around the lesion, following the shape of the demineralised area and ensuring that the patch falls upon sound enamel. The four crossed lines (iso-lines, iii) are used to alter the shape of the patch; c) when satisfied with the position and shape of the patch, the analysis is performed and the results shown. In this screen layout the original image, the grey scale reconstructed image and a lesion image is shown in addition to the quantitative data.

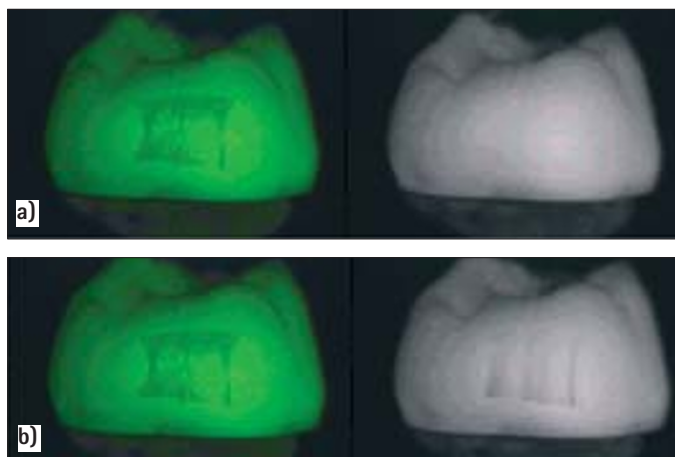


Fig. 2 a) ΔQ value -141.5 at 5% threshold. A well analysed image. Notice that the lesion is not visible on the grayscale reconstruction indicating that the analysis patch was well placed on sound enamel; and b) ΔQ value -70.2 at 5% threshold. A poorly analysed image. Notice the ghost-like presence of the lesion on the reconstructed image, indicating that the analysis patch crossed unsound enamel

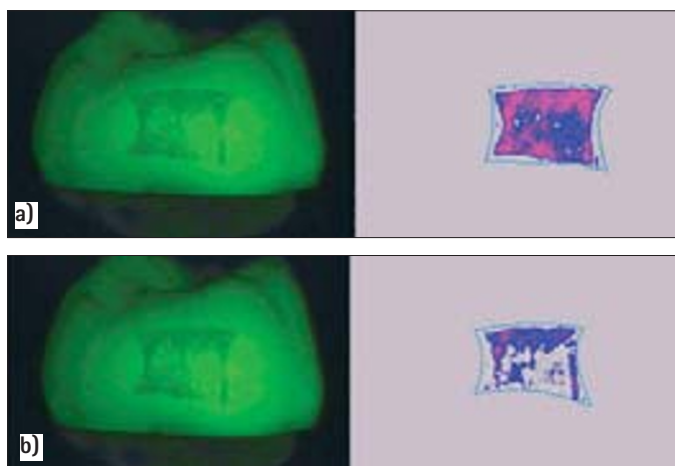


Fig. 3 a) A well analysed lesion. The map of the demineralised area (right) closely resembles that of the original image (left). This indicates that the analysis patch was correctly placed around the lesion and on sound enamel; and b) Poorly analysed lesion. The map of the demineralised area does not closely match the original lesion. It is likely that the analysis patch was placed on un-sound enamel

Final QLF images were taken of each lesion using a laboratory jack platform above which the QLF camera was held in a fixed position. Fine focus was achieved using the vertical control of the jack's platform. Using this method high quality QLF images were obtained. The images were stored on a PC and two further copies made of the image files to ensure exact replicas. Each set of 16 images was placed into a separate folder, and provided with a unique filename to ensure blinding of the examiners.

Examiners

The images were placed on to a CD-ROM and distributed to ten examiners at three different centres: The University of Liverpool, The University of Manchester and The University of Glasgow, all within the UK. Selection of the examiners was based upon achieving a spread of QLF experience. Accompanying the CD-ROM were examiner instructions and a proforma to elicit the following data:

- A quantitative assessment of each lesion's perceived difficulty of analysis (scale 1–5)
- A qualitative assessment of each lesion's perceived difficulty of analysis. Examiners were asked to describe why they allocated a particular rating to each lesion

The examiners were asked to complete this assessment on their first attempt only. The instructions included the following analysis 'rules' for examiners:

- Please use four iso-lines (to control the dimension of the patch).
- There is no need to use the contour markers (these are used to enable placement of an analysis patch on several images of the same tooth).
- Place the analysis patch around the lesion, following the shape as closely as possible, but ensuring that the border rests on sound enamel.
- Exclude a border if you feel that it is crossing 'un-sound' enamel.
- Following analysis, check to see if the reconstructed image is homogenous in appearance, with no white or black streaks and that the coloured reconstruction matches the outline of the original lesion.
- Repeat analysis if necessary to achieve the above.

These instructions were illustrated with pictorial examples of the process. The examiners were asked to leave a 'washout' period of 7 days between each of the three attempts. Following completion of the exercise the results were printed directly from the QLF program. The ΔQ value from each tooth and each attempt was recorded for statistical analysis.

Statistical techniques

Numerous statistical techniques for the analysis of reliability exist.⁵ For the data type presented here, ie continuous data, ANOVA and paired *t*-tests were used to detect significant differences between the values obtained by the examiners. In order to determine intra-examiner reliability, each of the three attempts was compared with each other using ANOVA and subsequent *t*-tests. When assessing the inter-examiner data, the mean of these data were used in a crosswise method with ANOVA and *t*-tests. The significance level was set at 95% ($\alpha = 0.05$). Raw data were entered into *Excel* and transferred to *SPSS v. 10* for statistical analysis.

RESULTS

Quantitative data

All ten examiners completed the study and returned correctly formatted data (100%). An example of an individual examiner's results is shown in Table 1. The results of the ANOVA and *t*-tests are shown in Table 2 for the intra-examiner assessment, and in Table 3 for the inter-examiner assessment. The mean difficulty rating for the 16 teeth was 2.5 ± 0.77 (Scale 1–5) demonstrating a good spread of analysis complexity.

Qualitative data

Each examiner was asked to describe the reasons that a particular complexity rating was allocated to each tooth. The comments made were consistent across all the examiners. Lesions classified as simple to analyse (1,2, or 3) were described as simple in shape and surrounded by clearly sound enamel. An example of such a lesion is tooth 14 of Figure 4. Complex ratings (4 or 5) were allocated to teeth that were described as uneven in shape and adjacent to areas of unsound enamel. Proximity to the EDJ or areas of shadow also led to a higher level of difficulty being analysed. When the analysis patches of teeth with ratings of 4 or above were examined, a large proportion had one or more borders excluded. Many examiners mentioned that this was a confusing aspect of the analysis. When looking at repeated analyses of the examiners, consistency of border inclusion/exclusion was low. Despite this, when the complexity of lesion scores were compared with variance scores, no relationship was found ($r = 0.4229$), suggesting that lesion complexity does not lead to a decrease in reliability.

Table 1 An example of the data set from a single examiner arranged according to Fleiss. Note the closeness of the means for each of the three attempts and the similar SD's (large as there is a spread of lesion severity and size, but small variation is the key factor). Rating is the value assigned by the examiner based upon lesion analysis complexity (1 easy, 5 difficult).

TOOTH	ATTEMPT 1	ATTEMPT 2	ATTEMPT 3	MEAN	SD	VARIANCE	RATING
1	-127.7	-104.5	-106.4	-112.9	12.9	165.9	1
2	-281.5	-272.7	-273.1	-275.8	5.0	24.7	4
3	-15.1	-14.1	-14.1	-14.4	1.0	0.3	1
4	-32.3	-32.4	-35.9	-33.5	2.1	4.2	2
5	-196.6	-174.6	-183.9	-185.0	11.0	122.0	2
6	-6.1	-7.8	-4.7	-6.2	1.6	2.4	3
7	-42.6	-37.4	-37.5	-39.2	3.0	8.8	5
8	-110.8	-103.4	-97	-103.7	7.0	47.7	2
9	-117.7	-122.4	-113	-117.7	4.7	22.1	3
10	-154.8	-162.9	-160.2	-159.3	4.1	17.0	1
11	-264	-296.8	-303.1	-288.0	21.0	440.7	4
12	-106.9	-110.8	-114.8	-110.8	4.0	15.6	5
13	-104	-103.2	-100.9	-102.7	2.0	2.6	3
14	-143.5	-142	-135.4	-140.3	4.3	18.6	1
15	-121.3	-127.3	-121.8	-123.5	3.3	11.1	2
16	-44.2	-47.9	-43	-45.0	2.6	6.5	3
Mean	-116.8	-116.2	-115.3	-116.1			
SD	80.9	83.3	84.9	82.8			Subject 1

DISCUSSION

The assessment of reliability is essential for any new scale, measuring device or group of calibrated examiners.⁶ The QLF can be regarded as a measuring device with two stages. This study has determined that the reliability or reproducibility of the analysis stage of the technique is good, both in terms of intra- and inter-examiner agreement. When assessing the intra-examiner data only one examiner (Examiner 4) was found to have significant differences between all three of their attempts. Two other examiners, (3 and 8) had a statistical difference between their first and last attempts. In total only 5 of 30 comparisons (16%) were significantly different and 3 (10%) of these were from a single, novice, examiner. The inter-examiner assessment demonstrated that only two differences were detected, both of which were between Examiner 4. Of 50 unique comparisons, only 1% created a significant difference as a result of an unreliable analysis. The precision of the device, ie the variability of repeated measures of a single phenomenon was high, resulting in few differences between and within examiner ratings.⁶

Examiner 4 was responsible for the most differences detected. This particular examiner was a novice at the technique, with very few hours of analysis experience. Many of the other examiners had either extensive experience with QLF or examining carious lesions using computerised analysis techniques such as transverse micro-

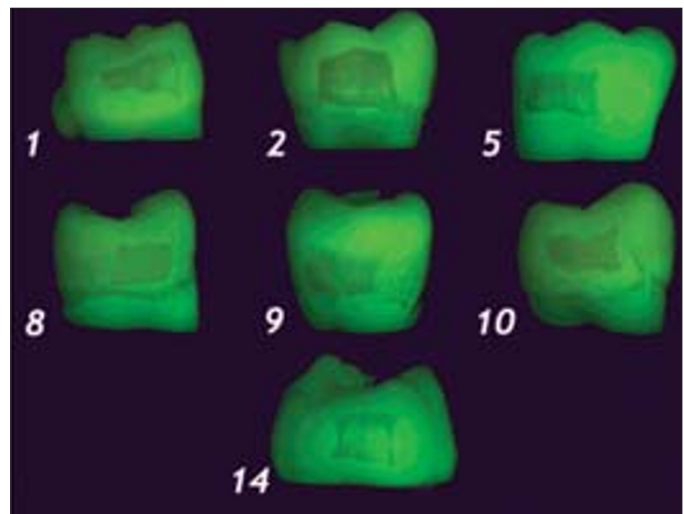


Fig. 4 Examples of the lesions presented to the examiners. Note that some lesions are surrounded entirely by clearly sound enamel (1,10,14), while others are closer to areas of shadow, gingival margin or other areas of 'non-sound' enamel. Lesion morphology ranges from plain rectangle or square (2, 14) to more complex shapes (1, 10)

Table 2 Intra-examiner reliability as assessed by significant differences (Bold numbers – significant ($\alpha = 0.05$), Non bold numbers – not significant)

EXAMINER	ATTEMPT 1 VERSUS ATTEMPT 2	ATTEMPT 1 VERSUS ATTEMPT 3	ATTEMPT 2 VERSUS ATTEMPT 3
1	$P = 0.427$	$P = 0.314$	$P = 0.217$
2	$P = 0.100$	$P = 0.444$	$P = 0.404$
3	$P = 0.165$	$P = 0.022$	$P = 0.069$
4	$P = 0.001$	$P = 0.007$	$P = 0.004$
5	$P = 0.104$	$P = 0.140$	$P = 0.361$
6	$P = 0.110$	$P = 0.150$	$P = 0.310$
7	$P = 0.350$	$P = 0.407$	$P = 0.278$
8	$P = 0.162$	$P = 0.004$	$P = 0.610$
9	$P = 0.360$	$P = 0.078$	$P = 0.085$
10	$P = 0.181$	$P = 0.089$	$P = 0.242$

radiography (TMR).⁷ While the TMR software operates in a different manner, familiarity with the computer processes is likely to be of value when transferring skills to QLF. Despite the difference detected between examiners 3 and 4's intra-examiner reliability, when the mean of these examiner's results were entered into the inter-examiner score, agreement was high, with no differences detected. This finding suggests that novices may find that repeating their analyses and reporting mean ΔQ values may be appropriate. It is well accepted that repeated measures could increase reliability and reduce any examiner or technique error. The training of novices by more experienced users is recommended, and, by using examples such as those shown in Figures 2 and 3, the ability to critically assess the analysis demonstrated. The employment of analysis rules will also facilitate new users' abilities to produce reliable results.

Lesions that lie close to un-sound enamel, enamel defects, stain or anatomical features, which are likely to confound the

Table 3 Inter-examiner reliability as assessed by significant differences (Bold numbers – significant ($\alpha = 0.05$), Non bold numbers – not significant)

Examiner	1	2	3	4	5	6	7	8	9	10
1	intra	$P = 0.437$	$P = 0.328$	$P = 0.261$	$P = 0.320$	$P = 0.243$	$P = 0.160$	$P = 0.432$	$P = 0.294$	$P = 0.266$
2		intra	$P = 0.178$	$P = 0.101$	$P = 0.206$	$P = 0.317$	$P = 0.218$	$P = 0.378$	$P = 0.417$	$P = 0.258$
3			intra	$P = 0.063$	$P = 0.200$	$P = 0.188$	$P = 0.245$	$P = 0.073$	$P = 0.107$	$P = 0.198$
4				intra	$P = 0.061$	$P = 0.069$	$P < 0.05$	$P = 0.231$	$P = 0.119$	$P < 0.05$
5					intra	$P = 0.414$	$P = 0.285$	$P = 0.231$	$P = 0.437$	$P = 0.463$
6						intra	$P = 0.142$	$P = 0.214$	$P = 0.391$	$P = 0.424$
7							intra	$P = 0.142$	$P = 0.127$	$P = 0.281$
8								intra	$P = 0.280$	$P = 0.178$
9									intra	$P = 0.441$
10										intra

analysis, are perceived as being more difficult to analyse. Despite this, the reliability of measurements of such lesions is still high. The provision of 'rules' within the instructions may permit 'calibration' of examiners in large studies. The authors recommend that QLF examiners should follow basic guidelines to ensure that research reports can be compared. In articles reporting the use of QLF the authors suggest that a description of any analysis is included, for example 'borders were excluded on un-sound enamel; those teeth requiring two or more border exclusions were removed from the trial'.

The QLF software is easy to master and the graphical representations of the original image, the reconstructed image and the lesion all enable users to determine if their analysis is appropriate. When using an analysis technique such as QLF, in which there is an element of subjectivity based upon examiner determinations, these guides are invaluable to both the novice and the expert.

CONCLUSIONS

This study has determined that, when using simulated, etched, *in vitro*, lesions the intra- and inter-examiner agreement of QLF analysis is high. In a multi-centred, multi-examiner trial, high levels of both intra- and inter-examiner agreement were measured. The data suggest that should significant differences be detected between QLF examinations then this can be attributed to the experimental variable rather than any examiner error. Novices in the technique may wish to perform repeated analyses

of their images to reduce measurement error. Experts in the technique may wish to compare their results to ensure that their techniques are producing reliable results. The use of analysis rules may help to increase reproducibility among and within examiner groups.

QLF has been shown to be a user-friendly, reproducible and reliable method of quantifying mineral loss from enamel. Further research is required to investigate if such high reliability is seen when analysing *in vivo* lesions, many of which may be more complex than those used in this trial. Lesions were considered complex when close to areas of un-sound enamel or areas that appeared un-sound on QLF.

1. al-Khateeb S, ten Cate J M, Angmar-Mansson B, *et al.* Quantification of formation and remineralization of artificial enamel lesions with a new portable fluorescence device. *Adv Dent Res* 1997; **11**: 502-506.
2. Ando M, Hall A F, Eckert G J, Schemehorn B R, Analoui M, Stookey G K. Relative ability of laser fluorescence techniques to quantitate early mineral loss *in vitro*. *Caries Res* 1997; **31**: 125-31.
3. Lagerweij M, van der Veen M, Ando M, Lukantsova L, Stookey G. The validity and repeatability of three light-induced fluorescence systems: an *in vitro* study. *Caries Res* 1999; **33**: 220-226.
4. Ross G. Caries diagnosis with the DIAGNOdent laser: a user's product evaluation. *Ont Dent* 1999; **76**: 21-24.
5. Fleiss J L. *The design and analysis of clinical experiments*. New York: John Wiley & Sons, 1986.
6. Brunette D M. *Critical thinking: understanding and evaluating dental research*. London: Quintessence Pub. Co, 1996.
7. Anderson P, Elliott J C. Scanning X-ray microradiographic study of the formation of caries-like lesions in synthetic apatite aggregates. *Caries Res* 1985; **19**: 403-406.

Submitting illustrations to the BDJ

Authors submitting manuscripts for publication in the BDJ are reminded that two copies of all illustrations must be supplied. This will assist in speeding up the refereeing process – manuscripts and illustrations are always sent to referees and a set of illustrations is required at the editorial office for reference. If you choose to submit 35-mm transparencies then one set of these and one set of prints (colour or black & white) will be acceptable.

We strongly advise authors to keep copies of all illustrations submitted in case letters are lost or damaged in transit. Please do not send glass-mounted transparencies – even when securely wrapped the glass will often arrive cracked or shattered in damage to the surface of the slide which usually renders it unsuitable for reproduction.