J Clin Periodontol 2004; 31: 815–819 doi: 10.1111/j.1600-051X.2004.00560.x
Printed in Denmark. All rights reserved

Journal of Clinical Periodontology

Accuracy and reproducibility of two manual periodontal probes

An in vitro study

Eralp Buduneli[1], Okay Aksoy[2], Timur Köse[3] and Gül Atilla[1]

[1]Department of Periodontology, School of Dentistry, Ege University, İzmir, Turkey; [2]Department of Mining Engineering, School of Engineering, Dokuz Eylül University, İzmir, Turkey; [3]Department of Computer Engineering, School of Engineering, Ege University, İzmir, Turkey

Buduneli E, Aksoy O, Köse T, Atilla G: Accuracy and reproducibility of two manual periodontal probes. An in vitro study. J Clin Periodontol 2004; 31: 815–819. doi: 10.1111/j.1600-051X.2004.00560.x. © Blackwell Munksgaard, 2004.

Abstract

Background: Periodontal probe is the standard instrument for assessment of probing pocket depth and clinical attachment level. However, probing measurements have inherent measurement errors from a wide range of sources, such as instrument, patient, examiner and disease status.

Aim: The purpose of the present study was to create an in vitro model simulating gingival sulcus/pocket and investigate the accuracy and reproducibility of two different manual probes.

Methods: Thirty-three aluminium blocks with dimensions of 2×2 cm and thickness ranging from 2.00 to 10.00 mm were constructed. Holes with a diameter of 1.00 mm were made in the blocks through the whole thickness. These 33 aluminium blocks were then carefully stucked together so that the probing faces would be on the same level. A plastic material, which shows deformation with forces >45 g was placed at the base of the blocks. Two conventional manual probes (Williams and WHO probes) were used in the present study. Seventeen periodontists were selected to perform duplicate measurements on the blocks over two visits using both of the probes. The intra- and inter-examiner percentage accuracy (with regard to 0.25 mm) and reproducibility for each of the duplicate measurements was calculated and analysed using repeated measure analysis of variance (ANOVA) (three factor experiments with repeated measure on two factors; probe and probing session).

Results: ANOVA showed statistically significant differences between the examiners ($p = 0.000$) and between the two readings of each probe ($p = 0.002$), while the differences between the two probes were not statistically significant ($p = 0.098$). The overall percentage of accuracy was higher with WHO probe compared with Williams probe. Kappa analysis revealed better reproducibility percentages for WHO probe in comparison with Williams probe.

Conclusions: This in vitro study, using a metal construction and a plastic material with a deformation coefficient similar to that of gingival pocket, may be suggested as a good model to test intra- and inter-examiner differences in periodontal probing. Our findings emphasise the importance of inter-examiner calibration for probing, particularly in longitudinal studies.

Key words: accuracy; deformation; in vitro; manual probe; probing; reproducibility

Accepted for publication 1 December 2003

Pocket probing has been and continues to be an important diagnostic tool to determine the presence and severity of periodontal diseases. In diagnosis and evaluation of treatment of periodontal diseases, probing pocket depth (PPD), probing attachment level (PAL) and bleeding on probing (BOP) are currently the most frequently used and the most informative parameters to estimate the severity of inflammation as well as the response to treatment (Caton et al. 1981). Furthermore, the current gold standard for assessing periodontal disease progression involves recording two periodontal attachment level measurements over an adequate time interval. Obtaining reliable measurements of pocket depth and attachment level is obviously critical to both longitudinal clinical studies and routine clinical assessment of periodontal therapy.

Evaluation of pocket depth and attachment level depends on the use of

the periodontal probe. The utmost importance of periodontal probing for diagnostic as well as therapeutic purposes has created considerable interest in the mechanisms and functions of the probe in order to improve its effect. Numerous studies have been published concerning the reliability of repeated measurements using various periodontal probes. The reproducibility data of periodontal probes in in vivo studies have been evaluated by describing the standard deviation of differences between replicate measurements of pocket depth. Conventional periodontal probes have been described as having a standard deviation of $\pm$ 0.82 mm (Gibbs et al. 1988).

The reproducibility of probe measurements has been shown to be dependent on a number of variables, which may be related to the pocket (Listgarten 1980), the clinician (Freed et al. 1983, Watts 1987) and the probe itself (Van der Velden 1978, Garnick & Silverstein 2000, Neto et al. 2001). These possible sources of error derived from the probe, angulation and disease status have been shown to affect the accurate measurement of the true pocket depth. Moreover, in case of repeated measurements by the same examiner or in cases where pocket depth measurements are performed by different examiners, accurate reproducibility of the measurement of the pocket depth may be influenced by these factors. This constitutes a major disadvantage of in vivo studies as it is not clear which of these factors is contributing to the variation. Furthermore, in vivo studies are seldom capable of assessing accuracy, as the true depth of the pocket is not known. Yet, there are very few studies evaluating the accuracy and reproducibility of various periodontal probes in in vitro models (Samuel et al. 1997). The aims of the present study were therefore two-fold: (1) to create an in vitro model simulating gingival sulcus/pocket and (2) to investigate the accuracy and reproducibility of two commonly used manual probes (Williams and WHO probes, Chicago, IL, USA) in an in vitro model.

Materials and Methods

Thirty-three aluminium blocks with dimensions of 2 $\times$ 2 cm and thickness ranging from 2.00 to 10.00 mm were constructed. The difference in the thickness of each block was 0.25 mm. Holes

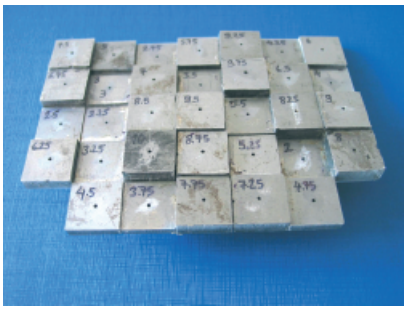

Fig. 1. Thirty-three aluminium blocks with thickness ranging from 2.00 to 10.00 mm were stucked together so that the probing faces would be on the same level.

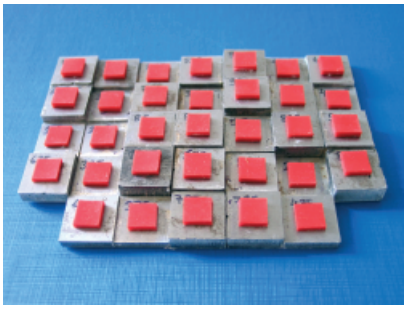

Fig. 2. A plastic material, which shows deformation with forces >45 g was placed at the base of the blocks.

1.00 mm in diameter were made in the blocks through the whole thickness. These 33 aluminium blocks were randomly allocated and then carefully stucked together so that the probing faces would be on the same level (Fig. 1). Various plastic materials were tested in Destructive Material Examination Laboratory of Dokuz Eylül University, School of Engineering, Department of Mining Engineering for their deformation characteristics and a plastic material to simulate the deformation properties of the bottom of gingival sulcus/pocket was sought. Amount of permanent deformation of these various plastic materials was designated by compressive test at constant load with Universal Shimadzu tensile equipment. The plastic material, which shows deformation with forces >45 g was selected and placed at the base of the blocks (Fig. 2). This material was replaced by new ones before each probing series. The construction formed with 33 aluminium blocks was then covered with plaster of Paris in a way that only the probing faces would be uncovered and the thickness of the blocks would not be seen (Fig. 3).

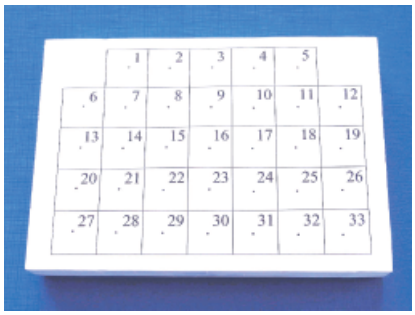

Fig. 3. This construction was then covered with plaster of Paris, leaving only the probing faces uncovered.

Two most commonly used manual probes; Williams and WHO probes were used in the present study. The Williams–Hu Friedy periodontal probe has a tapered tine with a tip diameter of 0.5 mm and the markings consist of darkly etched bands which divided the probe at 1, 2, 3, 5, 7, 8, 9 and 10 mm from the tip. The WHO–Hu Friedy periodontal probe has a ball at the tip with a diameter of 0.5 mm and markings at 3.5, 5.5, 8.5 and 11.5 mm from the tip.

Seventeen periodontists (nine males, eight females, 27–47 years of age) who have been in periodontal practice for at least 3 years were selected to perform duplicate measurements on the blocks over two visits using both of the probes. All measurements were carried out under standard conditions in terms of adequate light in the room. The examiners were neither aware of the true depths nor the range of depths of the holes being probed. The 33 holes were probed in random sequence by each examiner using both probes studied. In order to prevent recall of readings, duplicate measurements by each examiner using each periodontal probe were conducted with 3 days intervals.

Statistical analysis

The threshold limit of error in estimating the depth was selected as $\pm$ 0.5 mm for evaluation of accuracy. Thirty-three holes were probed with each probe and all probings were duplicated within 3 days.

The percentage of accuracy for each examiner was determined for each reading separately by counting the number of accurate (with regard to 0.5 mm) readings in a series of 33 readings and then this number was transformed to the percentage. The result of this calculation was considered as the

percentage of accuracy. This calculation method was performed for each probe.

To calculate the percentage of intra-examiner reproducibility, the number of reproduced readings was counted by comparing the first and second readings for each examiner and this number was transformed to percentage and considered as the percentage of reproducibility for each probe.

Then, this method of calculation was performed by comparing the first readings with Williams and WHO probe and counting the number of reproduced readings, which were considered as the reproducibility between the two different probes.

The reproducibility for each of the duplicate measurements was analysed using repeated measure analysis of variance (ANOVA) (three factor experiments with repeated measure on two factors; probe and probing session).

Furthermore, the reproducibility of different probes by each examiner were tested by kappa analysis for measure of agreement and the level of significance for all analyses was selected as $p < 0.05$.

Results

ANOVA showed statistically significant differences between the examiners ($p = 0.000$) and between the two readings of each probe ($p = 0.002$), whereas the differences between the two probes were not statistically significant ($p = 0.098$) (Table 1).

The interactions between the examiner and the probe was significant ($p = 0.000$), that is, some of the examiners exhibited better accuracy with Williams probe, whereas some did with WHO probe. The interaction between the examiner and the reading was significant ($p = 0.000$), that is some examiners measured accurately in the first reading, whereas some did in the second. The interaction between the probe and the reading was not significant ($p = 0.742$), that is, each probe showed similar accuracy data when the two readings were considered. The intra-examiner deviation was much smaller than the inter-examiner deviation.

The percentage of accuracy is described in Table 2. The overall percentage of accuracy was higher with WHO probe compared with Williams probe. Twelve of the 17 examiners estimated more accurately with WHO probe, while five examiners did with Williams probe.

Reproducibility percentages between readings 1 and 2 of two probes are shown in Table 3. When the reproducibility of each probe between readings 1 and 2 were compared, both of the probes showed highly reproducible data. The number of readings with statistically significant reproducibility was higher for WHO probe (16 examiners out of 17; 94.1%) in comparison with Williams probe (13 examiners out of 17; 76.4%) according to Kappa analysis ($p < 0.05$). When reproducibility percentages of each probe were considered, 13 of the 17 examiners exhibited higher percentages with WHO probe, whereas three examiners exhibited the opposite and one examiner showed the same reproducibility percentages for both probes. The frequency distribution of deviations between readings 1 and 2 of two probes are presented as percentages in Table 4.

When the reproducibility between Williams and WHO probes in the two readings (34 readings in total with each probe) was considered, the total number of readings with statistically significant reproducibility was 17, 10 of which were in readings 1 and 7 were in reading 2 ($p < 0.05$) (Table 5). The frequency distribution of deviations between two probes in readings 1 and 2 are presented as percentages in Table 6.

Discussion

The present study aimed at evaluating the accuracy and reproducibility of two

Table 1. Results of ANOVA

Source	Sum of squares	Degree of freedom	Mean square	*F*	*p*
examiner	22.24799	16	1.39050	12.602	0.000*
error (examiner)	56.49465	512	0.11034		
probe	0.61007	1	0.61007	2.905	0.098
error (probe)	6.72081	32	0.21003		
reading	1.02674	1	1.02674	11.535	0.002*
error (reading)	2.84826	32	0.08901		
examiner–probe	3.87099	16	0.24194	2.929	0.000*
error (examiner–probe)	42.29813	512	0.08261		
examiner–reading	6.43160	16	0.40197	6.494	0.000*
error (examiner–reading)	31.69340	512	0.06190		
probe–reading	0.00713	1	0.00713	0.110	0.742
error (probe–reading)	2.07375	32	0.06480		
examiner–probe–reading	1.99666	16	0.12479	1.631	0.057
error (examiner–probe–reading)	39.17246	512	0.07651		

*Significant difference ($p < 0.05$).

Table 2. The percentage of accuracy in readings 1 and 2 of two probes

Examiner	Williams probe reading 1	Williams probe reading 2	mean	WHO probe reading 1	WHO probe reading 2	mean
1	69.697	48.485	59.091	75.758	66.667	71.213
2	69.697	63.636	66.667	81.818	69.697	75.758
3	39.394	57.576	48.485	48.485	57.576	53.031
4	57.576	24.242	40.909	51.515	54.545	53.030
5	57.576	48.485	53.031	69.697	69.697	69.697
6	66.667	60.606	63.637	69.697	54.545	62.121
7	66.667	45.455	56.061	57.576	48.485	53.031
8	66.667	60.606	63.637	87.879	54.545	71.212
9	60.606	63.636	62.121	66.667	78.788	72.728
10	69.697	27.273	48.485	72.727	30.303	51.515
11	36.364	24.242	30.303	33.333	42.424	37.879
12	45.455	63.636	54.546	48.485	48.485	48.485
13	39.394	24.242	31.818	12.121	12.121	12.121
14	57.576	66.667	62.122	75.758	78.788	77.273
15	72.727	33.333	53.030	30.303	27.273	28.788
16	66.667	72.727	69.697	93.939	84.848	89.394
17	54.545	54.545	54.545	81.818	72.727	77.273
mean	58.645	49.376	54.011	62.210	55.971	59.091

Table 3. Reproducibility percentages between readings 1 and 2 of two probes

Examiner	Williams probe reading 1–reading 2			WHO probe reading 1–reading 2		
	Reproducibility (%)	κ	*p*	Reproducibility (%)	κ	*p*
1	66.6	0.341*	0.031	78.8	0.488*	0.004
2	75.7	0.457*	0.008	75.7	0.353*	0.032
3	69.7	0.413*	0.011	84.9	0.698*	0.000
4	66.6	0.382*	0.005	97.0	0.939*	0.000
5	72.7	0.457*	0.008	81.8	0.570*	0.001
6	75.7	0.478*	0.006	72.7	0.434*	0.009
7	72.7	0.471*	0.003	78.8	0.578*	0.001
8	75.7	0.478*	0.006	66.6	0.284*	0.019
9	72.7	0.421*	0.015	87.9	0.700*	0.000
10	51.5	0.178	0.142	57.6	0.280*	0.020
11	51.6	−0.128	0.443	84.8	0.681*	0.000
12	63.7	0.290	0.074	87.9	0.757*	0.000
13	72.7	0.388*	0.018	87.9	0.431*	0.013
14	72.7	0.426*	0.013	84.9	0.569*	0.001
15	48.5	0.105	0.407	72.8	0.336	0.053
16	75.8	0.429*	0.013	90.9	0.531*	0.001
17	87.9	0.756*	0.000	84.9	0.574*	0.001

*Significant reproducibility between readings 1 and 2 ($p<0.05$).

Table 4. Frequency distribution (%) of deviations between readings 1 and 2 of two probes

Examiner	Williams probe reading 1–reading 2					WHO probe reading 1–reading 2				
	⩾ −1	−0.50	0	+0.50	⩾ +1	⩾ −1	−0.50	0	+0.50	⩾ +1
1	3	6	49	30	12	3	–	67	21	9
2	–	3	64	33	–	–	3	70	18	9
3	18	–	52	–	30	9	–	76	–	15
4	–	6	49	39	6	–	3	97	–	–
5	3	3	64	24	6	–	12	76	12	–
6	24	18	55	–	3	3	3	67	21	6
7	3	–	55	33	9	3	3	79	15	–
8	24	18	55	–	3	–	3	61	24	12
9	3	9	73	15	–	3	12	79	6	–
10	–	6	34	39	21	–	–	30	49	21
11	–	21	34	36	9	6	15	73	3	3
12	24	18	52	3	3	12	–	82	–	6
13	3	3	67	12	15	6	18	58	18	–
14	12	15	55	15	3	–	18	73	9	–
15	3	3	24	52	18	3	21	46	24	6
16	6	15	58	15	6	–	12	79	9	–
17	–	15	76	9	–	9	9	58	21	3

Deviations of ⩾ −1, −0.50, 0, +0.50, ⩾ +1 mm.

commonly used manual probes in an in vitro model. The main advantage of such an in vitro study is the elimination of some major sources of error like the degree of inflammation in the periodontal pocket and angulation of periodontal probe. However, this fact limits the application of the results to the clinical situation. In the present study, a plastic material basing the holes was used to simulate the deformation of epithelial attachment at the base of the periodontal pocket. The aim was to approach to the clinical situation at least to some extent and to our knowledge this is the first study using such an in vitro model. The plastic material used in the present study showed deformation with forces >45 g and this limit is comparable with the recommended manual probing force.

Samuel et al. (1997) have published an in vitro study testing the accuracy and reproducibility of automated and conventional periodontal probes. In that study automated probes have been reported to offer increased accuracy over conventional probes and the reproducibility of both Florida pocket and disk probes has been found to be comparable with those of the conventional probes. They have used two aluminium blocks each having 15 holes and the various depths have been randomly allocated between the blocks. In the present study, we used 33 separate aluminium blocks with variable depths and stucked them together and covered with plaster of Paris in a way that the thickness of the blocks would not be seen. The holes with depths from 2.00 to 10.00 mm were distributed randomly in an effort to reduce bias. The only probe used in both studies is Williams probe. The results of the present study with regard to the accuracy and reproducibility of Williams probe are similar to the previous in vitro study (Samuel et al. 1997) in that the accuracy and reproducibility percentages are high. Accordingly, Williams probe is the most commonly used manual periodontal probe in clinical studies. In the present study, all the examiners were experienced periodontists whereas Samuel et al. (1997) have included two groups of examiners; experienced and inexperienced. They have reported statistically significant differences at the 5% level between the experienced and inexperienced examiners for the accuracy of the Williams probe. Surprisingly, a better performance was reported by inexperienced examiners. Furthermore, there was no statistically significant difference between experienced and inexperienced examiners with regard to the reproducibility of measurements with either probe. We consider that the more frequent the usage of the periodontal probe by the periodontist, the better accuracy and reproducibility can be achieved. According to our present findings, some periodontists exhibited better performance with the WHO probe, whereas some others did with the Williams probe. This finding may at least partly be explained by the familiarity or frequent usage of the probe by the examiner.

The present finding that the intra-examiner deviation is much smaller than the inter-examiner deviation reinforces the conviction that either a single examiner should do the follow-up probings particularly in longitudinal studies, or different examiners should be very well calibrated before taking part in the measurements. This fact has been stated also in clinical studies on accuracy and reproducibility of periodontal probes (Osborn et al. 1990, Grossi et al. 1996, Mayfield et al. 1996, Khocht & Chang 1998). In the present study, WHO probe exhibited better accuracy as well as better reproducibility percentages

Table 5. Reproducibility percentages between two probes in readings 1 and 2

Examiner	Reading 1 Williams probe–WHO probe			Reading 2 Williams probe–WHO probe		
	Reproducibility (%)	κ	*p*	Reproducibility (%)	κ	*p*
1	81.8	0.544*	0.002	63.6	0.280	0.085
2	75.7	0.353*	0.032	69.7	0.321	0.063
3	60.6	0.207	0.226	69.7	0.380*	0.029
4	63.7	0.269	0.119	51.6	0.074	0.604
5	69.7	0.355*	0.035	60.6	0.221	0.161
6	84.8	0.651*	0.000	69.7	0.382*	0.027
7	72.7	0.426*	0.013	66.7	0.331	0.056
8	72.7	0.270	0.059	63.6	0.258	0.135
9	75.7	0.478*	0.006	72.8	0.353*	0.030
10	84.8	0.631*	0.000	60.6	0.040	0.817
11	60.7	0.133	0.443	69.7	0.343*	0.032
12	72.7	0.453*	0.009	78.8	0.579*	0.000
13	60.6	0.061	0.643	75.8	0.205	0.200
14	69.7	0.343*	0.032	69.7	0.250	0.132
15	39.4	−0.028	0.817	63.6	0.143	0.407
16	66.6	0.057	0.606	81.8	0.468*	0.004
17	72.7	0.421*	0.003	69.7	0.368*	0.022

*Significant reproducibility between two probes ($p<0.05$).

Table 6. Frequency distribution (%) of deviations between two probes in readings 1 and 2

Examiner	Reading 1 Williams probe–WHO probe					Reading 2 Williams probe–WHO probe				
	⩾ −1	−0.50	0	+0.50	⩾ +1	⩾ −1	−0.50	0	+0.50	⩾ +1
1	–	21	61	12	6	12	18	49	12	9
2	3	9	58	21	9	–	18	52	15	15
3	15	–	39	–	46	3	–	54	–	43
4	3	15	55	21	6	9	33	46	9	3
5	3	18	61	15	3	3	39	46	12	–
6	6	9	73	9	3	–	3	52	15	30
7	3	9	58	24	6	–	33	52	12	3
8	9	18	67	3	3	–	3	49	21	27
9	6	15	64	12	3	3	24	64	9	–
10	3	18	67	12	–	6	24	46	15	9
11	3	21	40	27	9	3	33	58	6	–
12	6	9	67	6	12	6	6	58	9	21
13	–	9	55	18	18	6	12	58	15	9
14	15	9	58	12	6	3	24	55	15	3
15	3	6	6	61	24	9	12	40	24	15
16	–	15	43	36	6	–	12	55	24	9
17	12	36	43	6	3	9	30	55	6	–

Deviations of ⩾ −1, −0.50, 0, +0.50, ⩾ +1 mm.

compared with Williams probe. This finding is surprising, as Williams probe has more detailed markings, but may be explained by WHO probe's being easier to read. The design of WHO probe may be the underlying reason for this finding, since it has very clear black-and-white sections rather than just black lines. On the contrary, Williams probe has more detailed markings, which are just black lines. This difference in the design of both probes may have led to the differences between the accuracy and reproducibility percentage data. Therefore, the present results of κ analysis evaluating the reproducibility of duplicate readings emphasise the importance of consistency in the type of the periodontal probe used in clinical studies particularly in longitudinal and multicenter investigations.

As a conclusion, this in vitro study, comprising a plastic material with a deformation coefficient similar to that of the gingival pocket, may be a good model to test intra- and inter-examiner differences in periodontal probing. The present findings emphasise further the importance of inter-examiner calibration for probing particularly in longitudinal studies and such an in vitro model may be utilised for precalibration training prior to the enrolment of different examiners in a clinical study.

References

Caton, J., Greenstein, G. & Polson, A. (1981) Depth of periodontal probe penetration related to clinical and histologic signs of inflammation. *Journal of Periodontology* **52**, 626–629.

Freed, H. K., Gapper, R. L. & Kalwarf, K. L. (1983) Evaluation of periodontal probing forces. *Journal of Periodontology* **54**, 488–492.

Garnick, J. J. & Silverstein, L. (2000) Periodontal probing: probe tip diameter. *Journal of Periodontology* **71**, 96–103.

Gibbs, C. H., Hisrchfeld, J. W., Lee, J. G., Low, S. B., Magnusson, I., Thousend, R., Yerneni, P. & Clark, W. B. (1988) Description and clinical evaluation of a new computerized periodontal probe the Florida Probe. *Journal of Clinical Periodontology* **15**, 137–144.

Grossi, S. G., Dunford, R. G., Ho, A., Koch, G., Machtei, E. E. & Genco, R. J. (1996) Sources of error for periodontal probing measurements. *Journal of Periodontal Research* **31**, 330–336.

Khocht, A. & Chang, K.-M. (1998) Clinical evaluation of electronic and manual constant force probes. *Journal of Periodontology* **69**, 19–25.

Listgarten, M. A. (1980) Periodontal probing: what does it mean? *Journal of Clinical Periodontology* **13**, 165–176.

Mayfield, L., Bratthall, G. & Attström, R. (1996) Periodontal probe precision using 4 different periodontal probes. *Journal of Clinical Periodontology* **23**, 76–82.

Neto, J. B. C., Nogueira-Filho, G. R., Tramontina, V. A., Sallum, E. A., Nociti, F. H. & Sallum, A. W. (2001) Millimeter marks and probe tip diameter standardization from commercially available periodontal probes. A comparative study. *Journal of International Academy of Periodontology* **3**, 57–60.

Osborn, J., Stoltenberg, J., Huso, B., Aeppli, D. & Pihlstrom, B. (1990) Comparison of measurement variability using a standard and constant force periodontal probe. *Journal of Periodontology* **61**, 497–503.

Samuel, E. D., Griffiths, G. S. & Petrie, A. (1997) In vitro accuracy and reproducibility of automated and conventional periodontal probes. *Journal of Clinical Periodontology* **24**, 340–345.

Van der Velden, U. (1978) Errors in the assessment of pocket depths in vitro. *Journal of Clinical Periodontology* **5**, 182–187.

Watts, T. (1987) Constant force probing with and without a stent in untreated periodontal disease: the clinical reproducibility problem and possible sources of error. *Journal of Clinical Periodontology* **14**, 407–411.

Address:
Dr Eralp Buduneli
Department of Periodontology
Ege University
School of Dentistry
35100 Bornova, İzmir, Turkey
Fax: +90 232 388 03 25
E-mail: eralp@bornova.ege.edu.tr