



REPEATABILITY AND INTERDEVICE RELIABILITY OF TWO PORTABLE COLOR SELECTION DEVICES IN MATCHING AND MEASURING TOOTH COLOR

Panagiotis E. Lagouvardos, DDS, DrDent,^a Aggeliki G. Fougia, DDS,^b Sofia A. Diamantopoulou, DDS,^c and Gregory L. Polyzois, DDS, DrDent, MScD^d
Dental School, University of Athens, Athens, Greece

Statement of problem. There is a need to know how shade selection devices perform in matching and measuring tooth color, since these functions are usually evaluated independently and may present significant discrepancies.

Purpose. The purpose of this study was to evaluate 2 devices which offer both functions to test the null hypothesis that they present no differences in their repeatability and interdevice reliability relating to the 2 functions.

Material and methods. Thirty-one extracted anterior human teeth were measured twice, with each of the devices (ShadeEye NCC and VITA EasyShade), by one investigator experienced and calibrated with both devices. $L^*a^*b^*$ values and shade matches to VITA Classical and Vitapan 3D-Master shade guide systems were determined for all teeth. Paired t tests and intraclass correlation coefficients (ICCs) were used to statistically analyze the data ($\alpha=.05$).

Results. The results showed that $L^*a^*b^*$ values with VITA EasyShade were significantly higher than those with ShadeEye NCC ($P<.01$), with a mean difference of 10.01 ΔE units. Device repeatability ICCs for measuring tooth color ranged from 0.693 to 0.929, with significant differences between devices only with respect to parameter L^* ($P=.042$). Intradevice ICCs for matching tooth color ranged from 0.714 to 0.894, with significant differences between devices only with respect to the VITA Classical shade system ($P=.042$). Interdevice ICCs for measuring tooth color ranged from 0.402 to 0.486, with no significant differences between $L^*a^*b^*$ parameters, while interdevice matching reliability ranged from 0.189 to 0.430, with no significant differences between shade systems ($P>.05$).

Conclusions. All tooth color parameters were higher with the VITA EasyShade device. Measuring repeatability of the VITA EasyShade was higher than the ShadeEye NCC only for the L^* parameter. Matching repeatability of the ShadeEye NCC was higher than the VITA EasyShade for the VITA Classical system. Interdevice measuring reliabilities were not different for the color parameters, but matching reliability of the VITA Classical system was higher than that of the Vitapan 3D-Master. (J Prosthet Dent 2009;101:40-45)

CLINICAL IMPLICATIONS

Clinically, each of the evaluated devices is expected to have a high repeatability of the CIE $L^*a^*b^*$ measurements and the matching shades of a tooth color. However, they provide different $L^*a^*b^*$ values and shades for the same tooth. Lightness (L^*) and shade guide systems affect interdevice reliability.

Visual tooth shade selection is characterized by a high intraexaminer¹ and interexaminer variability,²⁻⁴ due to the numerous subjective factors that affect color selection. Electronic shade selection devices have the po-

tential for more accurate and reliable selection of a tooth color,^{5,6} since they are not influenced by the significant parameters in visual selection, such as illumination and operator variability.⁷ Matching validity and variability^{1,6,8}

seem to improve with the use of shade selection devices, and some investigators⁹ have suggested that these devices can be useful in determining color quality for clinical shade matching of porcelain crowns.

Presented at the 31st European Prosthodontic Association Conference, Athens, Greece, October 2007.

^aAssociate Professor, Department of Operative Dentistry.

^bPostgraduate student, Department of Operative Dentistry.

^cPostgraduate student, Department of Operative Dentistry

^dAssociate Professor, Division of Removable Prosthodontics.

The devices are colorimeters, spectrophotometers, spectroradiometers, image analyzers, or a combination of these. The most common types are the colorimeters and the spectrophotometers, which calculate the tooth color by measuring the amount and spectral composition of light reflected on the surface of the tooth. They usually express the results in the CIE $L^*a^*b^*$ system,¹⁰ as well as in one or more of the conventional shade guide systems. CIE $L^*a^*b^*$ is one of the standard color models used to describe all visible colors, using 3 basic coordinates. L^* represents the lightness of the color, a^* represents green (negative a^*) to red (positive a^*) colors, and b^* represents the blue (negative b^*) to yellow (positive b^*) color range. Spectrophotometers estimate the color by measuring the intensity of the reflected light in all visible wavelengths, while colorimeters measure the intensity of the reflected light filtered by red, green, and blue filters (simulating RGB absorbance).

These differences in the way intraoral shade matching instruments interpret the reflected light may directly affect the measuring functions of the devices, and indirectly, their matching functions. The matching function is based on the extrapolation of color measurement values into more meaningful values for a clinician, such as the shade tabs of a guide system. This presents a source of variation in the reliability of electronic and visual matching.^{4,11}

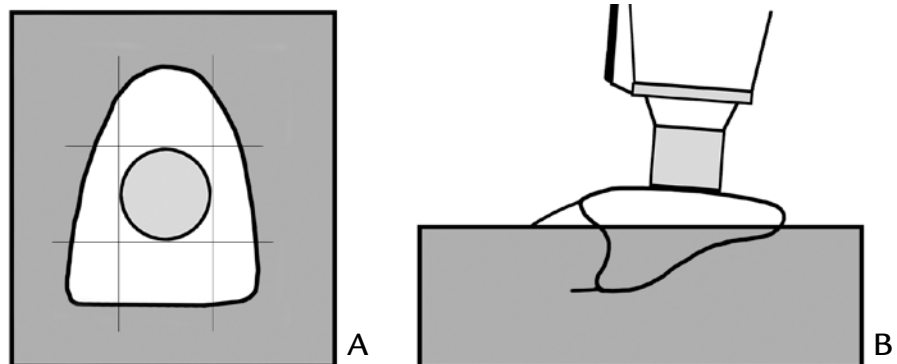
Two commercially available shade matching instruments which can be used for measuring as well for matching teeth color are the ShadeEye NCC (Shofu Dental, Kyoto, Japan) and the VITA EasyShade (VITA Zahnfabrik, Bad Säckingen, Germany). The first has been marketed since 1997, and the second has been available since 2002. Only 2 published studies related to these devices^{9,12-17} compare them,^{9,13} but neither compares their color measuring and matching functions as independent functions. These 2 functions, although interrelated, serve different purposes in clinical practice and may

not be consistent between the 2 devices. The measuring function indicates the exact position of a tooth color in the observer's standard color space, and it is useful for detailed tracking of color changes with time, as in tooth bleaching. The matching function indicates the degree to which a tooth matches a standard shade in a shade guide system and is useful for reconstructing the tooth substance with available tooth-colored restorative materials. Possible inconsistencies between devices in measuring color parameters or matching teeth to shade systems, as other studies have indicated,^{13,17} will result in difficulties in using the devices interchangeably, as well as in the communication of color between the clinician and technician, if they own different devices. According to ISO definitions,¹⁸ consistency of a measuring or matching instrument in repeated measurements expresses its repeatability, while the degree of agreement among measuring or matching devices expresses their interdevice reliability. Systematic errors are difficult to manage in these devices and affect instrument accuracy,¹⁹ meaning that interinstrument agreement is relatively poor in comparison with intrainstrument reliability.²⁰ For these reasons, the 2 devices were studied to test the null hypotheses that they present no differences in their color measuring and matching repeatability or interdevice reliability (interdevice agreement), as determined by the color parameters or shade systems.

MATERIAL AND METHODS

In this study, 31 recently extracted anterior human teeth, stored in tap water since the day of extraction, were measured with 2 color measuring devices, ShadeEye NCC (Shofu Dental) and VITA EasyShade (VITA Zahnfabrik). The teeth were selected to represent a wide range of shades, but with the least possible variation in surface morphology and convexity. For this reason, teeth with extremely convex, rough, or irregular surfaces were excluded. The specimens were measured twice with each instrument with an interval of 1 hour, at the middle third of their labial surface. Each tooth was stabilized in a horizontal position on a white silicone base (Rapid; Coltène/Whaledent AG, Altstätten, Switzerland) with its labial surface on top. Each tooth measurement was the mean of 3 consecutive measurements. During measurements, the instrument tip was placed perpendicular to and in complete contact with the labial surface of the tooth (Fig. 1). Two vertical and 2 horizontal lines on the silicone base helped to locate the position of the tip.

All measurements were made by a single investigator experienced with both devices, following the manufacturer's instructions. Before any measurement, the devices were calibrated on their own white ceramic block. The investigator was calibrated by repeated tests on standard shade tabs taken from the 2 guides, until he repeatedly



1 Tooth and device set-up during measurements. A, Labial view shows measurement area and guiding lines. B, Side view shows tooth and tip positioning.

demonstrated less than 1 ΔE unit between measurements and an ICC value greater than 0.85.

The color data was exported into the CIE $L^*a^*b^*$ system for the color measuring process, and into the VITA Classical (CL) (VITA Zahnfabrik) and Vitapan 3D-Master (3D) (VITA Zahnfabrik) shade guide systems for the color matching process. The devices express tooth shades in a shade guide system, using conversion scales for CIE $L^*a^*b^*$ values, based on internally stored data. This function is performed automatically in every measurement with VITA EasyShade (ES), but with ShadeEye NCC (SE), is a separate function. For this reason, another series of measurements had to be made with the SE device, to match the teeth to conventional shade guide systems. In this matching process, the SE device provides only a single match for the tooth, while ES gives a “strong” and a “second less possible” match for the tooth. In this study, only the “strong” match was considered with the ES device. It must be noted that when matching teeth to Vitapan 3D-Master (3D) shades with ES, an intermediate shade between 2 adjacent shades is often indicated. This interpolated intermediate shade (mixture of the 2 shades to achieve the intermediate) normally does not exist in the SE system, and for this reason, any intermediate shade was converted to the closest shade on the guide.

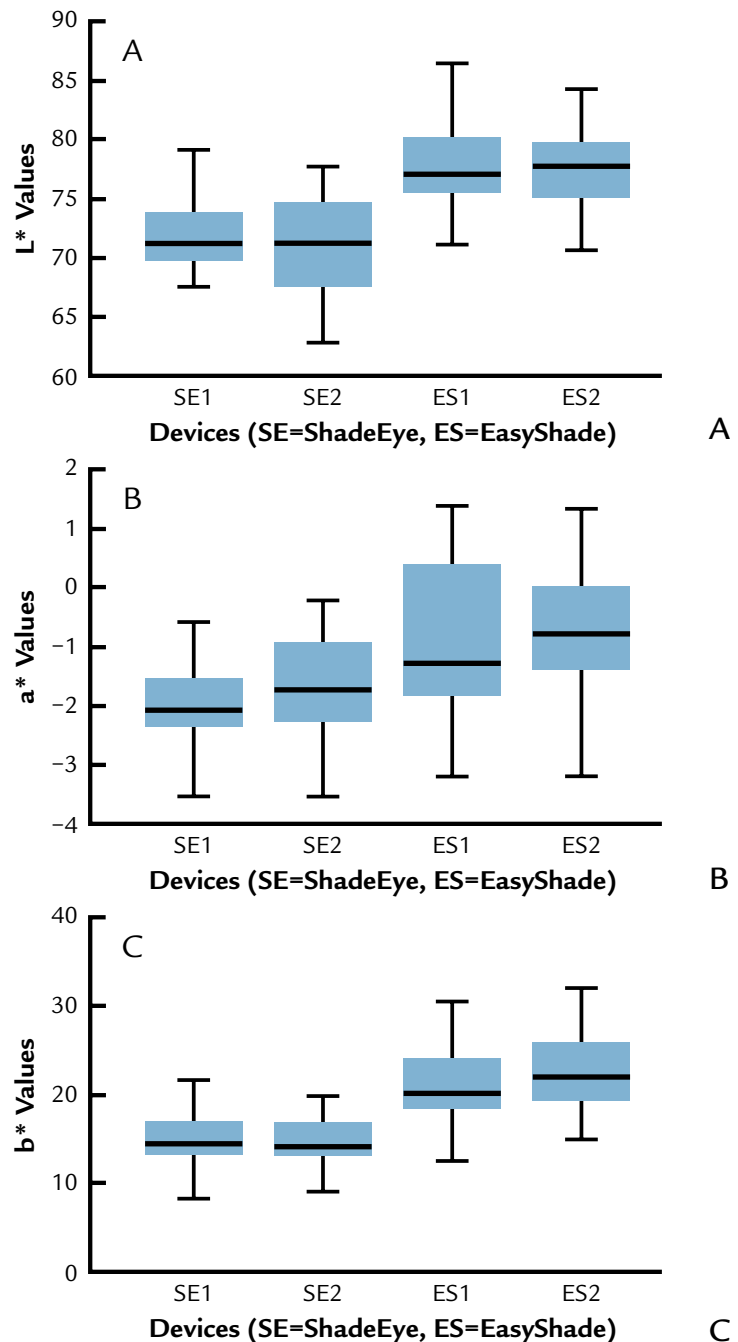
The level of device differences in measuring tooth color was evaluated using paired t tests on the actual $L^*a^*b^*$ values given by the devices. To estimate the repeatability and interdevice reliability of devices in measuring and matching tooth color, the intraclass correlation coefficient (ICC) was calculated as an average measure for absolute agreement of a 2-way mixed effect model. All tests were performed at an alpha of .05 with the use of statistical software (SPSS 10.0.7; SPSS, Inc, Chicago, Ill). Pairwise ICC comparisons were made using the Fischer z -transformation of the 2 coefficients, as suggested by Papoulis.²¹

RESULTS

The mean $L^*a^*b^*$ values of the teeth measured by the 2 devices are given graphically in the box-and-whisker plots in Figure 2. ES showed higher CIE $L^*a^*b^*$ values than SE for all color parameters. Paired t tests performed on these values showed significant differences among devices ($P<.001$) for all color parameters. However, no significant differences (except for b^* ,

with ES) were found between measurements within devices (Table I).

ICCs of devices based on $L^*a^*b^*$ measurements of teeth are shown in Table II, along with the differences between the horizontal pairs of coefficients. Intradevice ICCs (device repeatability) were slightly higher for ES than for SE, but statistical analysis showed a significant difference only for the L^* parameter. Analysis of interdevice ICCs (interdevice reliabil-



2 Box-and-whisker plots of CIE $L^*a^*b^*$ values of teeth with ShadeEye (SE) and EasyShade (ES) device for the first (1) and second (2) measurements. A, L^* values. B, a^* values. C, b^* values.

ity) showed moderate²² values for all color parameters, and no significant difference ($P=.921$) between them (Table II).

ICCs for repeatability and interdevice reliability in matching are shown in Table III, along with the P values between the horizontal pairs of coefficients. The higher ICC of SE, as compared to the ES device, was significant only for the CL system ($P_{CL}=.042$, $P_{3D}=.308$, respectively).

Agreement rates of devices in matching teeth shades to conven-

tional shade systems (CL and 3D) are given also in Table III and calculated for comparison with other studies. For an estimation of the overall agreement rate of the devices, both shades indicated by each device in the 2 subsequent matches were counted. The overall interdevice reliability was calculated in the same manner, for the CL and 3D systems, and the results showed significantly higher values for the CL than for the 3D system ($P=.036$) (Table III).

DISCUSSION

In this study, the results do not support rejection of the null hypothesis that measuring repeatabilities of the 2 devices were not different; however, only for the a^* and b^* parameters. The results also do not support rejection of the null hypothesis that matching repeatabilities of the devices were not different; however, only with the 3D shade system. In addition, the results do not support rejection of the null hypothesis that interdevice measuring and matching reliabilities were not different.

Measurements of tooth color, L^* , a^* , and b^* values, within the same device, were not different between the 2 measurements, except for b^* when it was measured with the ES device. The color difference of the measurements had an average value of 3.40 and 3.31 ΔE units for the SE and ES devices, respectively. These values remain below the 50% acceptability level of 5.5 ΔE units,¹⁵ but exceed the 50% perceptibility threshold for a clinical mismatch (2.6 ΔE units).¹⁵ Since the investigator was well experienced with both devices, a significant part of these values may be attributed to tooth variance in surface morphology, convexity, and, primarily, enamel thickness and opacity, even though the teeth had been carefully selected. Minor variations in instrument repositioning could also have contributed to discrepancies in instrument consistency. Human tooth substance is difficult to measure consistently due to its polychromatic nature and its varying translucency and surface texture.

Analysis of repeatability (intradvice reliability) in measuring tooth color showed that ICCs for L^* , a^* , and b^* parameters ranged from excellent to perfect (0.693 to 0.929), with no differences within the same device. Comparisons showed that the repeatability of ES was significantly higher than SE only for the L^* parameter. This means that in sequential measuring of tooth reflectance, the ES device produces more reliable results.

TABLE I. Paired t test values, and significance (2-tailed P) of differences between and within devices, in measuring CIE $L^*a^*b^*$ color parameters

	Within Devices		Between Devices
	ShadeEye NCC t Value/ P	VITA EasyShade t Value/ P	t Value/ P
L^*	1.1/.266	1.1/.265	15.1/<.001
a^*	1.8/.076	0.5/.638	5.5/<.001
b^*	1.1/.276	2.9/.006	12.6/<.001

TABLE II. Intraclass correlation coefficients (ICCs) indicating device repeatability and interdevice reliability in measuring tooth color, based on mean of 2 measurements

	Device Repeatability			Interdevice Reliability
	SE	ES	P^+	
L^*	0.802	0.928	.042	0.486
a^*	0.677	0.814	.307	0.463
b^*	0.792	0.852	.282	0.402

⁺Probability of difference between horizontal pairs of ICCs

TABLE III. Intraclass correlation coefficients for repeatability of devices and interdevice matching reliability to shade guide systems. Agreement rates of measurements within and among devices for CL and 3D shade systems are in parentheses

	CL ICC (%)	3D ICC (%)	P^+
SE repeatability	0.894 (45.2)	0.813 (41.9)	.255
ES repeatability	0.714 (45.2)	0.756 (41.9)	.731
Interdevice reliability*	0.866 (64.5)	0.640 (22.6)	.036

⁺Probability of difference between horizontal pairs of ICCs

*Overall estimation based on all measurements

Comparing the color of teeth measured by the 2 devices, all color parameters were significantly higher with the ES device (mean $\Delta L=6.5$, mean $\Delta a^*=1.2$, mean $\Delta b^*=6.9$). As a result, the mean ΔE of the 2 devices (10.01 ΔE units) was well above the level for a clinically unacceptable color match (5.5 ΔE units).²³ These differences are probably due to differences in light illumination affecting the way the 2 devices illuminate the tooth, differences in internal design and viewing geometry affecting the way the instruments collect and measure the reflectance on the tooth surface, and differences in the internal CPUs and algorithms affecting the way they estimate the color tristimulus values. Since all tristimulus values were greater with ES, the most obvious reason for this is that the device collects and measures higher reflectance values than SE for the entire spectrum, and primarily the yellow area. However, this study was not designed to determine accuracy of the devices, and for this reason, the higher values of ES do not mean that the device is more (or less) accurate than SE. Accuracy determination is important because instruments must provide measurements of the true values of color parameters, without substantial deviation from each other. Otherwise, human color vision variance is simply replaced by instrument variance. Interdevice reliability is also important. If a clinician uses one instrument and the dental laboratory technician another, it is very probable that the tooth shade will be replicated poorly in the restoration, if the instruments indicate different shades. Therefore, in clinical practice, using instruments that provide different $L^*a^*b^*$ values would have no consequence for the final color of the restoration, if the device's extrapolation algorithm was properly established to correctly indicate the standard color of a shade tab in a guide. Whether or not color selection devices use correct interpolation algorithms remains unclear. Interdevice reliabilities in measuring tooth color and based on the

mean of the 2 measurements ranged, within a moderate level, from 0.402 to 0.486. These low values seem to support the view of Joiner et al¹⁹ that interdevice agreement is expected to be lower than intradevice reliability, due to difficulties in controlling systematic error in these devices resulting from device differences.

Calculation of agreement rates based on the 2 matches of teeth to shade tab systems with each device (intradevice agreements) showed similar, moderate values for the 2 devices (45.2%, 45.2%) or the 2 shade systems (45.2%, 41.9%). These agreements were lower than those found by Tung et al¹⁶ (82%), who studied only a small number of teeth with shades primarily from the A group of the Lumin Vacuum system in vivo. Low agreements for the SE were also found by Li and Wang,¹⁴ and, for this reason, SE was considered responsible for the unreliable matching. Interdevice agreement showed different rates for the 2 shade systems, with the lower for the 3D system. Lower agreements for the 3D system were also found by Hugo et al,¹⁷ although they did not examine the same devices.

Repeatability of devices in matching, based on ICCs, showed lower values for ES than for the SE device (0.714 and 0.756 and 0.894 and 0.813 for CL and 3D systems, respectively), but these were significantly different only when the CL system was used ($P=.042$ for CL, $P=.579$ for 3D). This means that SE repeatability was better in matching teeth to the CL than to the 3D shade system. Since no difference was found in the repeatability of devices with each of the shade systems ($P=.255$ for SE and $P=.731$ for ES), the higher reliability of SE with the CL system probably results from a more accurate interpolation algorithm for CL than for the 3D system, which is relatively new. Therefore, in spite of their differences and the variability of shade systems, the devices manage to have the same repeatability in matching, with either system. This indicates that CPUs inside the devices, with

proper manipulation of matching data, can offer valuable assistance.

When both measurements (matches) were used in the estimation of an overall interdevice agreement, the rates were 64.5% for the CL and 22.6% for the 3D system. This difference in rates suggests that shade guides present a source of variation between devices. Reliability analysis also shows significantly different interdevice reliabilities for the 2 shade systems. Since this estimation is actually based on the agreement of either of the 2 possible shades indicated by the first device with either of the 2 of the second device, interdevice reliability based on fewer shades (for example, 1 to 1) is expected to have lower agreement rates. For this reason, the manufacturer's suggestion that "only one measurement is necessary for a correct matching, because of the linearity of the measurements" cannot be accepted, and a more appropriate practice would be to choose the most frequently appearing shade from 3 instead of 2 consecutive matches.

Although this study could have been done in vivo, control of the variability of tooth shade, surface morphology, convexity, and opacity is difficult. This, together with the large number of measurements needed (about 15 within a certain period of time for each tooth), and the better control possible for repeated measurements, favored an in vitro study. Whether the results can be extrapolated to the in vivo situation remains to be demonstrated by further clinical studies.

This study showed that measuring reliability is high for both devices; therefore, they can be used in clinical practice for sequential or comparative measurements of teeth. However, since the mean ΔE between repeated measurements was 3.3 to 3.4 units, the devices would probably be useful for comparisons only in situations where a ΔE of more than 3.4 units is expected. Matching reliability with CL and 3D systems is about the same in both devices, but if the CL matching

system is used, the SE device performs better. Using both devices in clinical practice might be confusing, since they will most probably indicate different shades for the same teeth, due to their low interdevice reliability. The study was unable to distinguish the best device for measuring and matching purposes. Since no external standards have yet been accepted, only studies with reference color can provide information about a better performing device or technique.

CONCLUSIONS

Within the limitations of this in vitro study, the following conclusions were drawn:

1. CIE $L^*a^*b^*$ color values of teeth measured with the ES device were greater than with the SE device. This color difference was at a clinically unacceptable esthetic level ($10.01 \Delta E$ units).

2. Measuring repeatability was excellent to perfect for both devices and for all color parameters, but significantly higher with ES for the L^* parameter (ICC of 0.9285) ($P=.042$).

3. Measuring interdevice reliability was moderate for all color parameters (ICCs were 0.404 to 0.4865), with no significant difference as compared to each other.

4. Matching repeatabilities of devices were excellent to perfect (ICCs were 0.714 to 0.894), with no significant differences with respect to shade guide system.

5. Matching interdevice reliability based on all measurements was excellent (ICCs were 0.640 to 0.866) but significantly different for the 2 shade systems ($P<.05$).

REFERENCES

- Okubo SR, Kanawati A, Richards MW, Childress S. Evaluation of visual and instrument shade matching. *J Prosthet Dent* 1998;80:642-8.
- Hammad IA. Intrarater repeatability of shade selections with two shade guides. *J Prosthet Dent* 2003;89:50-3.
- Jarad FD, Russell MD, Moss BW. The use of digital imaging for colour matching and communication in restorative dentistry. *Br Dent J* 2005;199:43-9.
- Klemetti E, Matela AM, Haag P, Kononen M. Shade selection performed by novice dental professionals and colorimeter. *J Oral Rehabil* 2006;33:31-5.
- Paul S, Peter A, Pietrobon N, Hämmerle CH. Visual and spectrophotometric shade analysis of human teeth. *J Dent Res* 2002;81:578-82.
- Paul SJ, Peter A, Rodoni L, Pietrobon N. Conventional visual vs spectrophotometric shade taking for porcelain-fused-to-metal crowns: a clinical comparison. *Int J Periodontics Restorative Dent* 2004;24:222-31.
- Dagg H, O'Connell B, Claffey N, Byrne D, Gorman C. The influence of some different factors on the accuracy of shade selection. *J Oral Rehabil* 2004;31:900-4.
- Wee AG, Kang EY, Johnston WM, Seghi RR. Evaluating color match of different porcelain shade-matching systems. *J Esthet Dent* 2000;12:271-80.
- Dancy WK, Yaman P, Dennison JB, O'Brien WJ, Razzoog ME. Color measurements as quality criteria for clinical shade matching of porcelain crowns. *J Esthet Restor Dent* 2003;15:114-21.
- CIE (Commission Internationale de l'Eclairage). Colorimetry - technical report. CIE Pub. No. 15, 3rd ed. Vienna: Bureau Central de la CIE; 2004.
- Hassel AJ, Koke U, Schmitter M, Beck J, Rammelsberg P. Clinical effect of different shade guide systems on the tooth shades of ceramic-veneered restorations. *Int J Prosthodont* 2005;18:422-6.
- Wee AG, Monaghan P, Johnston WM. Variation in color between intended matched shade and fabricated shade of dental porcelain. *J Prosthet Dent* 2002;87:657-66.
- Dozic A, Kleverlaan CJ, El-Zohairy A, Feilzer AJ, Khashayar G. Performance of five commercial available tooth color-measuring devices. *J Prosthodont* 2007;16:93-100.
- Li Q, Wang YN. Comparison of shade matching by visual observation and an intraoral dental colorimeter. *J Oral Rehabil* 2007;34:848-54.
- Meireles SS, Demarco FF, dos Santos Ida S, Dumith Sde C, Bona AD. Validation and reliability of visual assessment with a shade guide for tooth-color classification. *Oper Dent* 2008;33:121-6.
- Tung FF, Goldstein GR, Jang S, Hittelman E. The repeatability of an intraoral dental colorimeter. *J Prosthet Dent* 2002;88:585-90.
- Hugo B, Witzel T, Klaiber B. Comparison of in vivo visual and computer-aided tooth shade determination. *Clin Oral Investig* 2005;9:244-50.
- International Organization for Standardization. ISO 5725-1:1994. Accuracy (trueness and precision) of measurement methods and results --Part 1: general principles and definitions. Geneva: ISO; 1994. Available at: <http://www.iso.org/iso/store.htm>
- Joiner A, Hopkinson I, Deng Y, Westland S. A review of tooth colour and whiteness. *J Dent* 2008;36 Suppl 1:S2-7.
- Douglas RD. Precision of in vivo colorimetric assessments of teeth. *J Prosthet Dent* 1997;77:464-70.
- Papoulis A. Probability and statistics. Englewood Cliffs: Prentice-Hall; 1989. p. 296.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- Douglas RD, Steinhauer TJ, Wee AG. Intraoral determination of the tolerance of dentists for perceptibility and acceptability of shade mismatch. *J Prosthet Dent* 2007;97:200-8.

Corresponding author:

Dr Pan E. Lagouvardos
Department of Operative Dentistry
Dental School, University of Athens
2 Thivon St, Goudi
115 27 Athens
GREECE
Fax: +30 210 746 1162
E-mail: plagou@dent.uoa.gr

Copyright © 2009 by the Editorial Council for
The Journal of Prosthetic Dentistry.