Interexaminer Reliability in Clinical Measurement of L*C*h* Values of Anterior Teeth Using a Spectrophotometer

Alexander J. Hassel, Dr Med Dent[a]/Anne-Christiane Grossmann, Dr Med Dent[a]/
Marc Schmitter, PD Dr Med Dent[b]/Zibandeh Balke[a]/Anja M. Buzello, Dr Med Dent[a]

Purpose: The objective of this study was to investigate interexaminer reliability in the clinical measurement of the L*C*h* (lightness/value, chroma, hue) values of anterior teeth using a spectrophotometer (Vita Easyshade). ***Materials and Methods:*** The basic color of the maxillary right central incisors and canines of 23 subjects was spectrophotometrically determined by 4 clinicians and an experienced user (development manager) of the spectrophotometer. Also, to analyze the effect of different training with the instrument on interexaminer reliability, 2 of the clinicians were instructed in the use of the spectrophotometer by the experienced examiner, whereas the others instructed themselves by studying the operating manual. ***Results:*** Agreement between all examiners was acceptable to excellent (intraclass coefficient > 0.4). The mean value of the measured differences for the central incisors of all subjects for L* values was 5 (for C* = 3.8, h* = 2.7 degrees) and for canines, the mean L* was 4.5 (C* = 3, h* = 1.6 degrees). Results from comparison of the 2 different training methods were inconsistent. Agreement with the experienced examiner ranged from not acceptable (C* values for incisors of self-instructed examiners) to excellent. ***Conclusion:*** The distribution of the measurements of 1 subject could lead to deviations in color, probably with clinical impact. For canines, the measurements were at least equally reproducible (in some cases significantly more reproducible) compared to central incisors. Because of the small number of examiners and the inconsistent results, it was not possible to reach a definite conclusion about the effect of different training methods on interexaminer reliability. *Int J Prosthodont 2007;20:79–84.*

Visual color determination by comparison of tooth color with a standard (eg, commercially available shade guides) is the most frequently applied method of color assessment in dentistry.[1,2] This procedure is regarded as difficult to reproduce and highly subjective; variables that affect shade selection include external light conditions, metamerism, age, sex, fatigue of the eye, experience, and, probably, color blindness.[2–5] In contrast, the human eye is very efficient at detecting even small differences between the colors of adjacent objects.[2,6] In this context, it has been suggested that the reproducibility of shade selection should be improved, as this could be of clinical benefit.[2] Instrumental methods for determination of tooth color are objective and more rapid than visual shade matching.[1,4] Computer-assisted spectrophotometers and colorimeters generate mathematically comparable L*a*b* (lightness, red/green, yellow/blue) or L*C*h* (value, chroma, hue) values that quantify color.[4,7] The L*a*b* or L*C*h* values describe a specific location in the 3-dimensional color space, defined by the Commission International d'Eclairage (CIE) as an international standard in 1976. The CIE L*a*b* color space has a vertical axis that indicates relative lightness (L*) and is a continuous scale of gray shades: perfect black has an L* value of 0, whereas perfect white is characterized by the L* value 100. The 2 horizontal axes a* and b* represent levels of red (+a*) and green (–a*) and

[a]*Assistant Professor, Poliklinik für Zahnärztliche Prothetik, Heidelberg, Germany.*
[b]*Associate Professor, Poliklinik für Zahnärztliche Prothetik, Heidelberg, Germany.*

Correspondence to: *Dr Alexander J. Hassel, Poliklinik für Zahnärztliche Prothetik, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany. Fax: +49-6221-565371. E-mail: alexander.hassel@med.uni-heidelberg.de*

Fig 1 Positioning the probe tip.

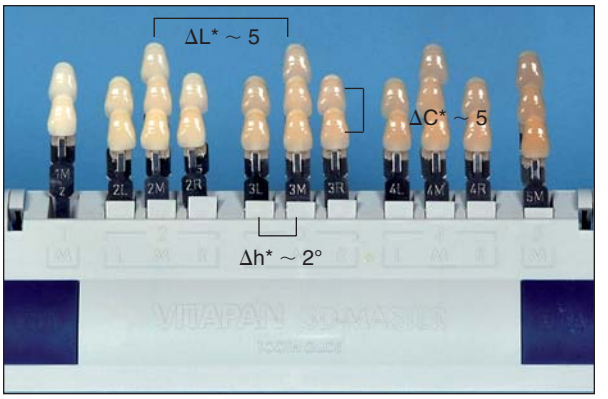

Fig 2 The Vita 3D-Master shade guide. The figure shows Δvalues for 2 groups with different lightness (ΔL*) and for 2 neighboring shade tabs in 1 lightness group (ΔC* and Δh*).

values of yellow (+b*) and blue (–b*). The CIE L*C*h* system enables representation of the CIE L*a*b* color space in cylindric coordinates. The L*C*h* system defines 3 visual aspects of color: L* (value) indicates the brightness of a color as a degree of lightness and darkness, C* (chroma) is the saturation of the color, and h* (hue) defines what is commonly called color, corresponding to the physical wavelength of light.[8]

It was stated in 1985 that a clinically applicable instrumental approach to the problem of dental color determination would be valuable.[7] This was still a problem in 1993, when a shade-matching instrument that gave reproducible results under clinical conditions was demanded.[9] In 1998 it was argued that spectrophotometric assessment of tooth color would provide more reproducible and accurate results than visual determination of L* values in vitro.[5] Other in vivo studies that examined tooth color visually and spectrophotometrically suggested that spectrophotometric shade determination is more accurate and reproducible than the conventional approach.[6,10] Studies comparing color changes of natural teeth in vivo have used spectrophotometric measurements for reference.[11]

The objective of this study was to measure the interexaminer reliability of color determination using a clinically applicable spectrophotometer (Vita Easyshade).

Materials and Methods

Spectrophotometer

The Vita Easyshade spectrophotometer was introduced to the dental market in January 2004. The instrument uses D65 illumination (6,500 K) for shade matching. Depending on the preset menu chosen, different modes of measurement are possible. For this study, the "Normal Mode Only" and "Tooth Single" settings were chosen to determine the basic shade of teeth.

Examiners

Four clinicians from the department of prosthodontics and a very experienced user of the Vita Easyshade—the development manager of the instrument (examiner EE)—served as examiners. Two of the clinicians (1 male, 1 female, called the "instructed examiners" [IE], IE1 and IE2) were instructed in the use of the instrument by EE. The instruction included general theory and handling of the instrument and a practical briefing. The positioning of the probe tip in the horizontal and vertical dimensions on the tooth surface for measuring basic tooth color was of particular importance. The IEs were instructed to place the probe tip above the region where most of the dentin is assumed to be; pictures of thin sagittal slices of different natural anterior teeth were used to identify this region. The other 2 examiners (1 male, 1 female, called the "self-instructed examiners" [SIE], SIE1 and SIE2) instructed themselves by studying the operating manual provided by the manufacturer. They maintained a minimum distance of 2 mm from the incisal edge and from the gingival tissue margin when placing the probe tip as close as possible to the tooth surface.

Participating Subjects and Investigative Procedure

Twenty-three employees of the department (11 men, 12 women, mean age 36.4 ± 8.7 years, range 23 to 50 years) volunteered to participate in this investigation. The objects of investigation were the maxillary right central incisors and canines; only unrestored teeth (n = 44) or teeth with minimal Class III restorations (n = 2) were included (because the probe tip had to be placed correctly without being in contact with the restoration). The basic color of the teeth was measured with a single measurement. Before measurement, an infection-control shield was accurately applied to the tip of the probe. The

Table 1 Mean Values, Ranges, Intraclass Correlation Coefficients (ICCs), and 95% Confidence Intervals (CIs) of the Measurements of All 5 Examiners

Tooth/parameter	Mean	95% CI	Range	ICC	95% CI lower bound of ICC
Central incisors					
L*	4.96	4.03/5.88	2–8	0.82	0.71
C*	3.78	2.94/4.63	0–7	0.73	0.59
h* (deg)	2.65*	2.19/3.12	1–5	0.99	0.99
Canines					
L*	4.48	3.79/5.17	1–7	0.83	0.73
C*	3.00	2.39/3.61	1–6	0.90	0.83
h* (deg)	1.61*	1.06/2.16	0–6	0.89	0.82

Significant difference between h values for incisors and for canines ($P < .001$).

instrument was calibrated after every subject (but not during measurements by different examiners on a single subject) using the calibration block supplied with the instrument. After the tooth had been wiped and dried with gauze (teeth were dried before every measurement), the probe tip was positioned (Fig 1) and the measurement button pressed. The test persons were examined consecutively by EE and then by the 4 other examiners in succession. Between examinations, subjects closed their mouths so that the teeth would not dry out. All examinations were conducted in the same location under natural light conditions.

Statistical Evaluation

Analysis of interexaminer reliability was performed by 2 methods. First, the measurements of all 5 examiners were compared. The ranges of differences between measurements of L*, C*, and h* values in each subject by all examiners were calculated. To clarify the clinical impact of these differences, the percentage of L*, C*, and h* ranges above $\Delta L^* = 5$, $\Delta C^* = 5$, and $\Delta h^* = 2$ degrees were reported. These limits were chosen in accordance with the shade tab arrangement of the Vita 3D-Master, a systematically arranged shade guide (Fig 2). According to the manufacturer, for 2 groups with different lightness $\Delta L^* \approx 5$, ΔC (chroma) and Δh (hue) between 2 neighboring shade tabs of one lightness group on the Vita 3D-Master are given by $\Delta C^* \approx 5$ and $\Delta h^* \approx 2$, respectively. In addition to descriptive analysis, the reliability of the L*, C*, and h* values was calculated using the intraclass correlation coefficient (ICC), a reliability coefficient (1 would be indicative of perfect agreement, 0 of no agreement; negative values are theoretical). Values of the lower bound of the ICC 95% confidence interval (95% CI) of $0.4 < ICC \leq 0.75$ were judged as acceptable and $ICC > 0.75$ as excellent.[12–14]

In a second analysis of interexaminer reliability, the L*, C*, and h* values from EEs were used as reference measurements. Descriptive analysis and ICC values were calculated for measurements of the IEs and SIEs in comparison with these reference measurements.

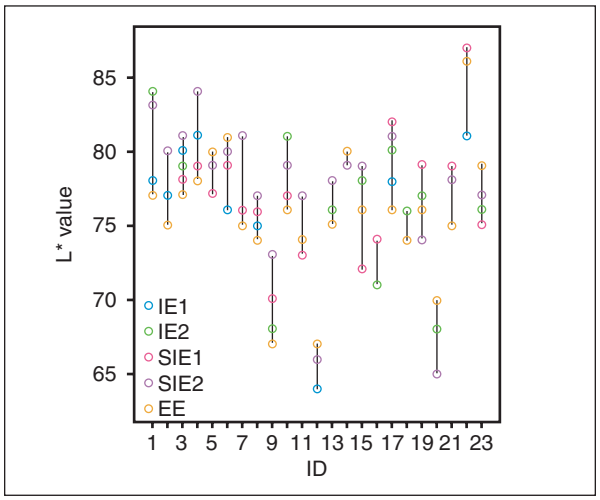

Fig 3 L* values given by each examiner for canines (overlapping measurements possible).

Because some variables exhibited a nonnormal distribution, the nonparametric Mann-Whitney *U* test was used to compare the 2 groups. Differences between the IEs and the SIEs (the influence of the single examiner was disregarded) and values for canines compared with central incisors were assessed. The independence of the groups was assumed. The level of probability of significance was set at $P < .05$. All statistical analyses were performed with SPSS Version 13.0.1 (SPSS Inc).

Results

Interexaminer Reliability for All Examiners

For the central incisors, agreement between examiners was acceptable for L* and C* values and excellent for h* values (an ICC value of 1 is caused by rounding) (Table 1). For canines, agreement was excellent for C* and h* values and acceptable for L* values. The range of differences between L*, C*, and h* values for all examiners on 1 subject varied from 0 (h* value) to a maximum of 8 (L* value). Figure 3 shows the ranges of L* values for canines. The means for L*, C*, and h* values

Table 2 Mean Δ (Absolute Values), 95% CI of Mean Δ, ICCs, and 95% CI Lower Bound (LB) of ICC for Each Examiner Versus the Measurements by EE of L*C*h* Values of Central Incisors (I) and Canines (C)

Examiner	Mean Δ	95% CI Δ	ICC	95% CI (LB)
IE1				
L* value (I)	2.78	1.97/3.60	0.81	0.60
C* value (I)	1.52*	0.90/2.15	0.83	0.63
h* value (I) (deg)	1.70†	1.11/2.29	0.99	0.98
L* value (C)	1.91	1.19/2.64	0.83	0.64
C* value (C)	1.48	0.88/2.07	0.88	0.73
h* value (C) (deg)	0.61†	0.32/0.89	0.94	0.86
IE2				
L* value (I)	1.83	1.15/2.50	0.92	0.81
C* value (I)	1.30*	0.55/2.06	0.76	0.51
h* value (I) (deg)	1.09	0.68/1.50	1.00	0.99
L* value (C)	1.87	1.13/2.61	0.86	0.70
C* value (C)	1.43	0.81/2.06	0.88	0.73
h* value (C) (deg)	0.83	0.40/1.25	0.87	0.71
SIE1				
L* value (I)	2.26	1.61/2.90	0.83	0.63
C* value (I)	2.00*	1.17/2.83	0.68	0.38
h* value (I) (deg)	1.35	0.87/1.83	0.99	0.99
L* value (C)	2.09	1.23/2.94	0.80	0.59
C* value (C)	1.13	0.63/1.62	0.92	0.82
h* value (C) (deg)	0.74	0.41/1.06	0.92	0.83
SIE2				
L* value (I)	2.65	1.67/3.63	0.75	0.50
C* value (I)	2.22*	1.49/2.94	0.65	0.33
h* value (I) (deg)	1.57‡	1.05/2.08	0.99	0.98
L* value (C)	3.00	2.11/3.89	0.77	0.52
C* value (C)	1.70	1.27/2.12	0.91	0.80
h* value (C) (deg)	0.78‡	0.41/1.15	0.89	0.76

*Differences between EE and SIE measurements were significantly greater than those between EE and IE measurements ($P = .043$).
†Differences between h* values for canines were significantly lower than those for central incisors ($P = .004$) for IE1.
‡Differences between h* values for canines were significantly lower than those for central incisors ($P = .02$) for SIE2.

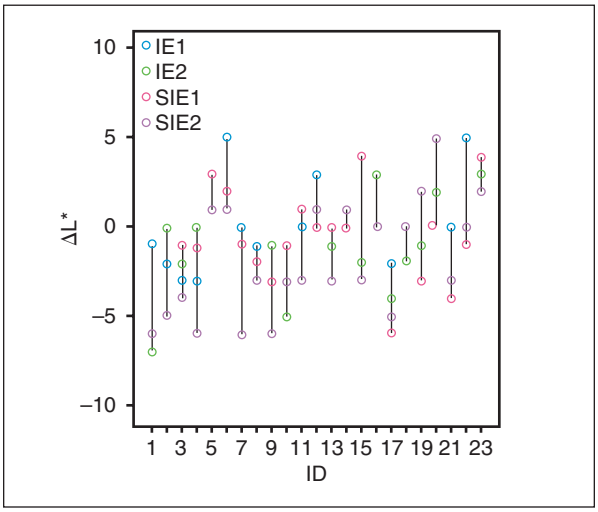

Fig 4 ΔL* values given by each examiner for canines; ΔL* = 0 represents the value measured by the EE (overlapping measurements possible).

were lower for canines, but this was significant only for h* values ($P < .001$). For 48% of incisors and 30% of canines, the range of measured differences was greater than $\Delta L^* = 5$; for 22% of incisors and 9% of canines, the range was greater than $\Delta C^* = 5$; and for 61% of incisors and 18% of canines, the range was greater than $\Delta h^* = 2$ degrees. Again, the values for canines were lower, but this was significant only for the h* values.

Comparison of Instructed Examiners with the Experienced Examiner

With the exception of the ICC of the C* values obtained for the incisors by the SIEs, all values for interexaminer reliability were judged as acceptable or excellent (Table 2). The mean differences (which could be positive or negative; absolute values are presented here) between results from IEs and those from the SIEs were not significant for either canines or incisors, except for the C* value for central incisors ($P = .043$, greater difference for SIEs). Mean differences between the values from all 4 examiners and those from the EE in this sample were lower for nearly all measurements of

Table 3 Percentage of Measurement Differences Versus EE Greater Than $\Delta L^* = 5$, $\Delta C^* = 5$, and $\Delta h^* = 2$ for IE and SIE Combined and for Each Group

	Central incisors			Canines		
	$\Delta L^* > 5$	$\Delta C^* > 5$	$\Delta h^* > 2$	$\Delta L^* > 5$	$\Delta C^* > 5$	$\Delta h^* > 2$
IEs and SIEs	7.6	4.4*	19.6†	7.6	0*	2.2†
IEs	6.5	2.2	21.7	2.2	0	2.2
SIEs	8.7	6.5	17.4	13	0	2.2

The percentage of measurements of C values greater than $\Delta C^* = 5$ was significantly higher for central incisors ($P = .044$) compared to canines.
†The percentage of measurements of h* values greater than $\Delta h^* = 2$ was significantly higher for central incisors ($P < .001$) compared to canines.

canines, but only for h* values from IE1 and SIE2 did this reach significance ($P = .004$ and $P = .019$, respectively). Figure 4 shows examples of measurement differences in the L* values for canines. Table 3 shows the percentage of measurement differences versus EE that were larger than $\Delta L^* = 5$, $\Delta C^* = 5$, and $\Delta h^* = 2$ degrees. There was no significant difference between IEs and SIEs with regard to the percentage of outliers. There was a significantly lower percentage of outliers for the C* values ($P = .044$) and for the h* values ($P < .001$) for canines versus central incisors.

Discussion

Because the material of human teeth is not homogeneous, it is not possible to determine the "true" basic color of a natural tooth, either with an electronic aid or by manual discrimination. The objective of this study was not, therefore, to investigate whether or not the color measured by the spectrophotometer was correct but whether it was possible to obtain reproducible measurements from different examiners. It was assumed that some of the disagreement would be caused not by the instrument but by the selection of different spots by the different examiners. It has already been established for colorimetric measurements that the spectral reflectance of a contoured surface cannot be duplicated unless the exact same spot is measured.[15,16] It is, indeed, difficult to determine which measurement from different examiners most reliably detects basic color; it is possible that all measurements are incorrect. Two different methods of analysis were therefore used. One calculated the interexaminer reliability of all 5 examiners, and the second determined the reliability of the measurements compared with a reference measurement. The measurement of the development manager of the Easyshade was defined as the reference value, because it was assumed he was an expert in handling the instrument. The L*C*h* values were chosen as the objects of investigation, because they define color mathematically and this definition of color assigns measurement errors to the different axes of tooth color. This approach is more informative than comparing the measured values with global shade tab colors. L*C*h* values rather than L*a*b* values were used, because the clinical impact of a measurement difference could be made clear more readily by reference to the systematically arranged Vita 3D-Master shade guide, which is based on the L*C*h* system with defined differences between the shade tabs. A measurement error of $\Delta L^* = 5$ would be the same ΔL^* seen between 2 groups of lightness on the shade guide, which is considered to be discernible by most human eyes.

Although the reliability of measurements among all examiners was acceptable or excellent, it could be demonstrated that the mean range of measurement differences was of clinical relevance; for example, for L* values, the mean was approximately 5 for central incisors and 4.5 for canines, and the maximum was 8. In addition, in 9% to 61% of the teeth, the ranges were greater than the difference between 2 neighboring shade tabs on the 3D-Master. On numerous occasions, therefore, different examiners would have chosen different shades. When measurements by the EE were used as references, these outliers, which made up 0% to 22% of measurements by the other examiners, were fewer. This uncertainty in reliability had to be considered when measuring basic color with a single measurement (as recommended by the manufacturer). Further research should investigate whether the use of a mean of multiple measurements would increase reliability. The ΔE value is often used to express differences between 2 measured colors. There are, however, differing opinions about which ΔE values—ranging from 0.4 with the highly trained human eye under laboratory conditions[9] to an average of 3.7 (rated as a match for compared teeth within the mouth)[17]—can be discriminated by the human eye and/or which values are acceptable. Johnston and Kao[17] found that the acceptability of color differences depended on patient-bounded factors: a color difference correlated to ΔE between 2.2 and 4.4 was acceptable, and ΔE between 3.8 and 9.3 was not. However, an unweighted ΔE does not distinguish between the different impacts of L*, a*,

and b* values on tooth color. It has been reported that observers are more sensitive and critical to color changes in the red range than changes in the yellow range, which correspond to identical values of ΔE.[18]

The interexaminer reliability of the SIEs compared with that of the IEs was inconsistent. Most analyses have found no differences between these groups, but there was a significantly larger measurement difference between C* values for incisors and unacceptable interexaminer agreement for the SIEs. There are, therefore, indications that "self-instruction" by reading the manual and becoming familiar with the instrument could be as effective as more complex introduction with practical briefing by an experienced examiner. This conclusion should, however, be treated with caution, because of the inconsistency of the results and the small number of examiners.

The measurement differences tended to be smaller for canines than for central incisors. This finding is remarkable, bearing in mind the convex shape of the canine and problems with correct positioning of the uncapped probe on the tooth surface.[11] It was expected that the flatter surface of the central incisor might be less prone to errors when the probe was applied. This result could be because the translucency of the central incisor is greater than that of the canine, which is supposed to increase the measurement error of the spectrophotometer.

Other studies of reliability have demonstrated that reproducibility with an intraoral colorimeter was 82%, whereas in visual color determination, reproducibility was 73%.[16] These results are indicative of high reproducibility, whereas Culpepper reported only 22% intraexaminer reproducibility for visual shade-taking.[3] Douglas achieved interexaminer reproducibility of color differences for 2 examiners of $\Delta E = 0.13$ or $\Delta E = 0.61$ when using a colorimeter in combination with an individually constructed positioning device that guarantees an identical measurement point.[19] Unacceptable reproducibility, ranging from $\Delta E = 1.1$ to $\Delta E = 32.1$, was achieved by use of a specially designed intraoral colorimeter; in vitro testing of this instrument furnished more reliable values.[9]

Conclusions

Differences between measurements, either among all examiners without a reference measurement or in comparison to a reference measurement, could lead to deviations of shade, probably with clinical impact (on the basis of the shade tab intervals on the Vita 3D-Master). Measurements of canines were as reliable or in some cases more reliable than those of central incisors, in contrast with expectations based on the convex shape of the canine and, therefore, problems positioning the probe tip. Interexaminer reliability after self-instruction with the manual was in most cases no worse than that after more profound instruction. Because of the small number of examiners and the inconsistent results, however, it was not possible to reach a definite conclusion about the effect of different training on interexaminer reliability.

Acknowledgments

We thank Vita Zahnfabrik for providing the Vita Easyshade. Special thanks to Mr Andre Schöne for serving as the experienced examiner.

References

1. van der Burgt TP, ten Bosch JJ, Borsboom PCF, Kortsmit WJPM. A comparison of new and conventional methods for quantification of tooth color. J Prosthet Dent 1990;63:155–162.
2. Hammad I. Intrarater repeatability of shade selection with two shade guides. J Prosthet Dent 2003;89:50–53.
3. Culpepper W. A comparative study of shade-matching procedures. J Prosthet Dent 1970;24:166–173.
4. Okubo SR, Kanawati A, Richards MW, Childress S. Evaluation of visual and instrument shade matching. J Prosthet Dent 1998;80:642–648.
5. Horn DJ, Bulan-Brady J, Lamar Hicks M. Sphere spectrophotometer versus human evaluation of tooth shade. J Endod 1998;24:786–790.
6. 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;24:222–231.
7. Goodkind RJ, Keenan KM, Schwabacher WB. A comparison of Chromascan and spectrophotometric color measurements of 100 natural teeth. J Prosthet Dent 1985;53:105–109.
8. Joiner A. Tooth colour: A review of the literature. J Dent 2004;32:3–12.
9. Goldstein GR, Schmitt GW. Repeatability of a specially designed intraoral colorimeter. J Prosthet Dent 1993;69:616–619.
10. Paul S, Peter A, Pietrobon N, Hammerle CHF. Visual and spectrophotometric shade analysis of human teeth. J Dent Res 2002;81:578–582.
11. Russell MD, Gulfraz M, Moss BW. In vivo measurement of colour changes in natural teeth. J Oral Rehabil 2000;27:786–792.
12. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med 1990;20:337–340.
13. Bland JM, Altman DG. Measurement error and correlation coefficients. Br Med J 1996;313:41–42.
14. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. Comput Biol Med 1989;19:61–70.
15. Seghi RR. Effects of instrument-measuring geometry on colorimetric assessments of dental porcelains. J Dent Res 1990;69:1180–1183.
16. Tung FF, Goldstein GR, Jang S, Hittelmann E. The repeatability of an intraoral dental colorimeter. J Prosthet Dent 2002;88:585–590.
17. Johnston W, Kao E. Assessment of appearance match by visual observation and clinical colorimetry. J Dent Res 1989;68:819–822.
18. Douglas RD, Brewer JD. Acceptability of shade differences in metal ceramic crowns. J Prosthet Dent 1998;79:254–260.
19. Douglas RD. Precision of in vivo colorimetric assessments of teeth. J Prosthet Dent 1997;77:464–470.