BJOS
Brazilian Journal of Oral Sciences

Original Article

Volume 21
2022
e225946

Reproducibility and comparison between methods for gingival color evaluation: a validation study

Matheus Völz Cardoso[1,*], Giovanna Vicenzoti[1], Adriana Campos Passanezi Sant'Ana[1], Mariana Schutzer Ragghianti Zangrando[1], Carla Andreotti Damante[1]

[1] Periodontics, Bauru School of Dentistry, University of São Paulo.

Aim: This study aims to evaluate and validate the sensibility and the level of agreement between different gingival color measures obtained by a spectrophotometer (SPECTRO) and a photography (PHOTO) method. **Methods:** Among 40 patients, the color was measured 2 mm apical to the gingival margin by CIE L*, a*, b* system using a reflectance spectrophotometer and the photography's plus software. The level of agreement between three different measures (m1, m2, m3) in parameters L*, a*, b*, and ΔE (color variation) was evaluated by random and systematic errors, as well as the limits and coefficient of concordance. A comparison between the methods was performed by the Bland-Altman test and the sensibility level was evaluated accordingly to the ΔE: 3.7 thresholds with $p<0.05$ as the level of significance for these comparisons. **Results:** The SPECTRO method has not presented the systematic error ($p>0.05$) and had reproducibly and agreement level in three variable measures L* (r: 0.6), a* (r: 0.3), and b* (r: 0.5) as to the PHOTO method L* (r: 0.6), a* (r: 0.5), and b* (r: 0.5), which presented systematic error in L* values ($p<0.05$). The means of ΔE between measurements were: 6.5 SPECTRO and 5.9 PHOTO. There was no good level of sensitivity ΔE> 3.7 and agreement between the methods, mainly for the a* values. On the other hand, for the L* and in for the most comparisons of b* values, the level of agreement was higher. **Conclusion:** Both methods could quantify the gingival color from the coordinates L *, a *, and b *, which has shown greater reliability between the measurements acquired by the SPECTRO method.

Keywords: Color. Gingiva. Photography, dental.

Corresponding author:
Matheus Völz Cardoso
Bauru School of Dentistry
Alameda Octávio Pinheiro Brisolla 9-75
17012-901 Bauru-SP -
Phone: 55 14 32358278 -
Fax: 55 14 32234679 -
Cell Phone 55 14 98112-2293
matheusvolz@alumni.usp.br

Editor: Dr. Altair A. Del Bel Cury

Received: June 6, 2022
Accepted: January 17, 2022

Introduction

Color is a complex phenomenon and an interesting field of study in several areas of science. Complexity is derived from the physical, physiological, geometric, and sensorial color presentation characteristics, and this conditions bring complications for the evaluation and categorization of this phenomenon[1]. Color has a wide utility in dentistry. Color perception and selection are a daily maneuver in dental office attendance. Several factors can modify the dental[2] and gingival color perception, such as visual organ fatigue, changing backgrounds, or differences in light incidence. For the understanding of the subjectivity of color perception in science, the *Commission Internationale de l'Eclairage* (CIE) has established different tridimensional color spaces. These measurement systems aim to upgrade global communication and to exclude the major subjective factor of color analyses. The L⋆ a⋆ b⋆ color space had three fundamental components, illumination: L⋆, variations between red/green: a⋆, and yellow/ blue: b⋆. The CIElab color space had values with major correlation compared to human perception[3]. Through tridimensional color spaces, it is possible to calculate and compare objects, specimens, and periods[4], using the color variation (ΔE).

Due to scientific advances, the color of dental enamel can be evaluated[5,6] with a higher level of evidence compared to gingiva color analysis[4]. The primary methods of color acquisition were derived from colorimeters and spectrophotometers devices. Under specific adjustments, digital cameras showed potential as an alternative method to color evaluation[7]. When dental enamel is evaluated, digital cameras associated with software were accurate tools to evaluate L⋆ and b⋆, but not for the analysis of the a⋆ axis[8,9]. Studies to the gingival color evaluation have tried to establish gingival color shade guides[3,10-12], as exhibited by dental enamel[6]. However, in order to classify the gingival color, the main variations are race/ethnicity and gender[13,14]. Besides that, other not-yet evaluated factor, are the possible interference of different gingival phenotypes[15] and had an importance in the validation of gingival color methods. Despite that, the gingival color was studied with these tools after the gingival graft procedures[16], tissue changes derived by different colors and materials of dental implant abutments[17], differences between natural and artificial gingival colors[18], also reduction in gingival inflammation (redness) after periodontal treatment against gingivitis[19]. Photographs were used to compare the ΔE threshold color values among different evaluators. This study has reported that professionals and patients had a distinct sense of color perception and different ΔE thresholds[20].

Comparisons between color analysis methods for dental enamel have shown that photography and colorimeters have obtained ΔE values below of recognized by the human eye[21,22]. They have also presented reliable sensitivity, due to the error of the analysis of the measurement for both methods, which were below of 2 units in the central incisors[21]. In clinical conditions, the accurate gingival color evaluation is achievable to compare results in randomized clinical trials (RCTs) before and after periodontal plastic surgery procedures (gingival grafts and aesthetic crown lengthening), in implant dentistry, as described before[16,17] to compare different

implant abutment materials, the gingiva luminescence, and to explain for a patient the acquired soft tissue color match. In the future, it will be possible the patient conducts a self-evaluation of the gingival inflammation levels by photographs. These methods can reduce the subjective analyses currently executed in RCTs, highly dependent on patients and on professional individual criteria examination. In addition, objective color parameters could be used as a complementary result, not a substitute, in gingival aesthetic evaluation executed by the patient. This would facilitate the gingival color comparison between different populations, ethnicities, and cultural realities of the participants of the study.

Despite that, the comparison amongst methods for gingival color evaluation and the measurements errors were poorly evaluated and validated in literature. When one method is established, the level of reproducibility is important to the validation of the color acquisition protocols tested. This study has aimed to evaluate the sensibility (agreement, reproducibility, and assertive grades) for three different measurements (m1, m2, and m3) for each method of gingival color evaluation by spectrophotometer (SPECTRO) and photography (PHOTO), and to compare both methods. The null hypothesis (1) is that the level of agreement between each measure (m1, m2, m3)/ intra-method shows equivalent results, and the second hypothesis (2) is that both methods present similar values in the L*, a*, b*, and ΔE outcomes.

Material and methods

Forty patients were invited to participate in this observational study that evaluated the gingival color in the same period by two the SPECTRO and PHOTO methods. The local human ethics committee from Bauru School of Dentistry approved the study protocol (nº 2.505.538 /CAAE: 79080117.4.0000.5417). The inclusion criteria were: 1) patients with teeth without signals of disease activity (periodontitis or gingivitis), 2) full month bleeding and plaque index ≤ 20%; 3) probing depth < 3 mm on teeth; 4) no restored and/or endodontically treated teeth; 5) aligned teeth with arch and adjacent teeth; 6) health and intact alveolar mucosa (without irritation signals, burn injuries or other lesions); 7) nonsmoker patients; 8) at least twice a day brush frequency.

The exclusion criteria were: 1) smokers; 2) systemic diseases that affect the cicatricial course or blood dyscrasias; 3) using medications such as anticonvulsant, antihypertensive, also cyclosporine; 4) Pregnant woman; 5) patients with notable cutaneous alteration due to excessive tanning and/or manifestations of skin diseases; 6) patients already submitted to a periodontal surgical procedure on the included site; 7) teeth (crown or root) with color alteration due to endodontic lesions; 8) remarkable alterations in alveolar or keratinized mucosa (pigmentations; trauma; amalgam tattoo, and melanic pigmentation).

The reflectance spectrophotometer[11,12,17] (direct color acquisition method) and the photography plus software (indirect color acquisition method)[16,19,20] were used methods for the gingival color evaluation. The measurements were performed in triplicate by the same evaluator (GV). The reproducibility of methods was evaluated in each patient by three different measurements (m1, m2, m3) at the same site both for the SPECTRO and the PHOTO methods.

The measures in the SPECTRO group were performed by a reflectance spectrophotometer for dental color analysis Easy shade (VITA). The selected area was in the center of a buccal site, 2 mm apical to the gingival margin. The equipment was protected by a plastic film and was calibrated as determined by the manufacturer. The range of the spectrum was set between 400 to 700 nm, and it was programmed to generate values of L*, a*, and b* axes.

In the PHOTO group, a photograph was taken in the same area evaluated by the spectrophotometer. A standardized photography protocol was established based on previous studies[9,19,23]. The digital camera (T6 model, Canon do Brasil Indústria e Comércio) with a magnification ratio of 1:1 was selected in the 100 mm macro lens (Canon do Brasil Indústria e Comércio). It was also used a macro ring flash (Canon do Brasil Indústria e Comércio). The camera focus was adjusted manually. The equipment configuration was standardized at ISO 100 and diaphragm aperture of 32, using 4500 Kelvin as color temperature. The same operator (MVC) was responsible for obtaining all the images, ensuring the acquisition process. The color was measured in the center of a buccal site 2 mm apical to gingival margin in each tooth, and the size of program cursor was adjusted to dropper>color sorter tool in the software (Adobe Photoshop CS6®)[19]. All the measures were performed by the same operator (GV) on the same screen computer (Sony VAIO®) with standardized screen settings[24]. The software was configured to generate values of L*, a*, and b* axes.

In the tested methods, the patient was positioned in a comfortable chair, with the head positioned in the headrest of a dental chair. It was used lip retractors on both sides of the mouth (Maquira, Maringa, Paraná- Brazil) to promote access of to the anterior upper teeth, which was included in study. The external conditions of color measures were set under natural clinical room light always during the day in a period of 10:00 a.m. to 4:00 p.m. For each method, the ΔE was calculated by the equation 1, which was based on previous methodologies[16,20,25].

Equation 1:

$$\Delta E = \sqrt{(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2}$$

Sample size

It was used a previous study[26], which evaluated the measurement error, in order to calculate the sample size of this study. It was determined that the minimum number of pairs of comparisons that needed to be executed was of 25-30 pairs[27] to evaluate the random error. 20 pairs of comparisons were the minimum number to evaluate the Dahlberg error method[28]. The sample size was estimated in forty patients and sites, and the null hypotheses assumed a comparison power >80% (β: 0.83) determined by software GPower 3.1.

Statistical analyses

The data distribution was tested by the Shapiro-Wilk test (n<50). For the reproducibility analyses, L*, a*, and b* values were compared using the three different measurements of mean and median to each method separately, using the Wilcoxon or the

T-tests and the Pearson or Spearman correlation tests[29,30]. In order to calculate the Dahlberg coefficient (measurement error)[31], the concordance limits, and the standard error of the method sensitivity, it was evaluated the dental enamel studies, considering the value of ΔE <3.7[21,22], which is under the threshold of the human eye, considered as adequate sensitivity.

To evaluate the second null hypothesis, for the comparison between the SPECTRO versus the PHOTO method, on which the values of L*, a*, b*, and ΔE (direct-indirect) were equal to zero[5], were considered to be the perfect concordance. The Bland-Altman analysis was performed for the comparison between methods in each measurement, besides the 2D graphic comparisons presentation with the confidence interval[32,33]. All analyses adopted the significance level of p<0.05 and the data was analyzed in the statistical software (IBM SPSS Statistics)[34] and RProject[35].

Results

From the patients (n: 40) enrolled in this study (n: 10 men and n: 30 women), the average age of the group was of 39.4 years (maximum 57 and minimum 20 years). All the participants evaluated themselves as of the white race-ethnicity. The gingival sites evaluated were located on the upper right central incisor (n: 18/45%) and on the upper left central incisor (n: 22/55%). The values obtained in the three different measures for any method were presented in Table 1. These values were calculated using the ΔE equation for each measure and method (Table 1). Overall, the ΔE values were similar for each method, with ΔE: 6.5 for the SPECTRO and 5.9 for the PHOTO. Comparing these three measurements, all of them have surpassed the reference value (ΔE: 3.7), which is already considered perceptible by the human eye. The level of reproduction between the measures was evaluated in three comparison levels (m1 versus m2, m1 versus m3, and m2 versus m3) (Table 2). Mainly, in the PHOTO method, the values of the L* axis in two comparisons surpassed the significance value (p<0.05).

Table 1. Values of L*a*b*and ΔE between measures (m1, m2, and m3) for SPECTRO and PHOTO methods and (n: 40).

SPECTRO	M1_ L	a	b	M2 _L	a	b	M3 _L	a	b	ΔE M1 vs M2	ΔE M1 vs M3	ΔE M2 vs M3
mean ± sd (max;min)	57.3±8.6 (77.2; 40.3)	18.2±10.3 (58; 6.2)	15.2±11.0 (61; 4.5)	58.4±7.7 (74.7; 37.5)	17.8±5.8 (38; 11.3)	14.3±6.2 (39; 9.6)	59.4±9.2 (72.4; 35)	17.3±6.6 (37.3;10.2)	15.0±9.3 (58.8; 9.1)	6.7 ± 4.5 (23; 0)	7.2 ±5.2 (19.8; 0)	5.6 ± 4.3 (15.9; 0)
Median, 25%,75%	57.1	17	12.7	57.1	16.7	12.7	60.6	16.4	12.8	5.5;	5.8;	5.3;
	51.8	13.8	11.4	53.4	14.2	11.1	57.2	12.9	11.6	4;	5.3;	2;
	63.5	20.1	13.8	63.8	19.3	14.2	65.8	19.0	14.8	8.6	10.4	8
PHOTO												
mean ± sd (max;min)	59.3±6.1 (77; 47)	33.1±5.5 (47; 21)	18.2±3.3 (28; 11)	59.8±6.2 (75; 46)	33.6±5.3 (46; 20)	18.8±3.3 (26; 11)	57.4±5.5 (70; 47)	35.4±5.0 (44; 26)	19.6±3.6 (31; 11)	3.9 ± 3.8 (15.4; 0)	6.7 ± 4.5 (22; 0)	7.1 ± 4.0 (17.2; 0)
Median, 25%,75%	59.5	33	19	59.0	34	18	57	35	20	3;	5.9;	6.3;
	56	30	17	56.3	31	16	53.7	31	17	0;	3.5;	4.6;
	62.5	36.5	20	63.8	37.7	20	61	38	21	6.2	8.4	8.9

max.: maximum; min.: minimum

Table 2. Reproducibility between measures.

(n: 40)	Concordance limits	Concordance coefficient*	Dahlberg casual error	Mean of differences	standard deviation	Standard error	paired t test ↓	Pearson(r)↓ coefficient correlation	ICC
M1 vs M2 L *spectro*	-1.158	-11.5 – 9.19	3.77	-1.15	5.28	1.31	0.19	0.80	0.79
a	-0.86	-10.4 – 8.75	3.47	-0.86	4.91	1.05	0.29	0.31	0.30
b	-0.20	-6.99 – 6.58	2.42	-0. 20	3.46	0.36	0.89#	0. 5#	0.34
L *photo*	-0.13	-6.99 – 6.71	2.44	-0. 13	3.49	0.23	0.81	0.83	0.83
a	-0.25	-6.22 – 5.72	2.13	- 0. 25	3.04	0.49	0.62	0.83	0.83
b	0.02	-5.68 – 5.73	2.03	0.02	2.91	0.05	0.98#	0.5#	0.64
M1 vs M3 L *Spectro*	-1.25	-14.2 – 11.7	4.70	-1.25	6.62	1.13	0.26	0.69	0.67
a	-0.42	-8.83 – 7.98	3.0	-0.42	4.28	0.59	0.59	0.55	0.47
b	0.32	-7.48 – 8.14	2.79	0.32	3.98	0.49	0.59#	0.5#	0.03
L *Photo*	1.72	-7.05 – 10.4	3.35	1.72	4.47	2.30	**<0. 05**	0.72	0.69
a	-0.91	-11.3 – 9.4	3.75	-0.91	5.30	1.03	0.30	0.56	0.56
b	0.38	-7.05 – 7.83	2.66	0.38	3.79	0.61	0.69#	0.5#	0.49
M2 vs M3 L *Spectro*	-0.09	-11.7 – 11.5	4.13	-0.09	5.92	0.09	0.92	0.71	0.71
a	-0.44	-5.23 – 6.12	2.04	0.44	2.89	0.91	0.36	0.68	0.67
b	0.53	-4.72 – 5.79	1.91	0.53	2.68	1.19	0.33#	0.5#	0.44
L *Photo*	1.72	-7.05 – 10.4	3.35	1.72	4.47	2.30	**<0. 05**	0.72	0.69
a	-0.91	-11.3 – 9.4	3.75	-0.91	5.30	1.03	0.30	0.56	0.56
b	0.38	-7.05 – 7.83	2.66	0.38	3.79	0.61	0.69#	0.5#	0.49

* 95% confidence interval; ↓p values; # b axis values show nonparametric distribution (Shapiro-Wilk test. p<0.05): Wilcoxon test and Spearman correlation was executed; ICC: intraclass correlation coefficient.

The obtained concordance limits for the SPECTRO were L*: ≈1, a*: 0.4-0.8, and b*: 0.2-0.5. As for the concordance coefficients (closest values for maximum agreement), they have exhibited intervals of L*: 25, a*: 15, and b*: 12 units. For the PHOTO method, obtained the concordance limits were L*: 0.1-1.8, a*: 0.2-0.9, and b*: 0.3, and the concordance coefficients were L*: 15.3, a*: 16 e b*: 13 units. Predominantly in both methods, the values have shown reduced concordance limits and standard error (next to 1 and <1, respectively). The random error was 3-4 units to L* and a* axes, and 2 to b* axis, with 3 units (1.8 to -1.2) of variations between means. The similarity between measures is satisfactory (ICC: intra-class correlation coefficient: 0.44 SPECTRO/0.57 PHOTO) in both comparisons.

The results for the comparisons between the measurements to each method were presented in Figures 1 and 2 (Fig. 1 spectrophotometer, Fig. 2 photography). The values of the SPECTRO method showed major proximity to zero (value of maximal concordance between measures) and had shortest confidence intervals (ellipses), except for the L* axis. In the PHOTO method, a major level of differences was observed with values that move away from zero.

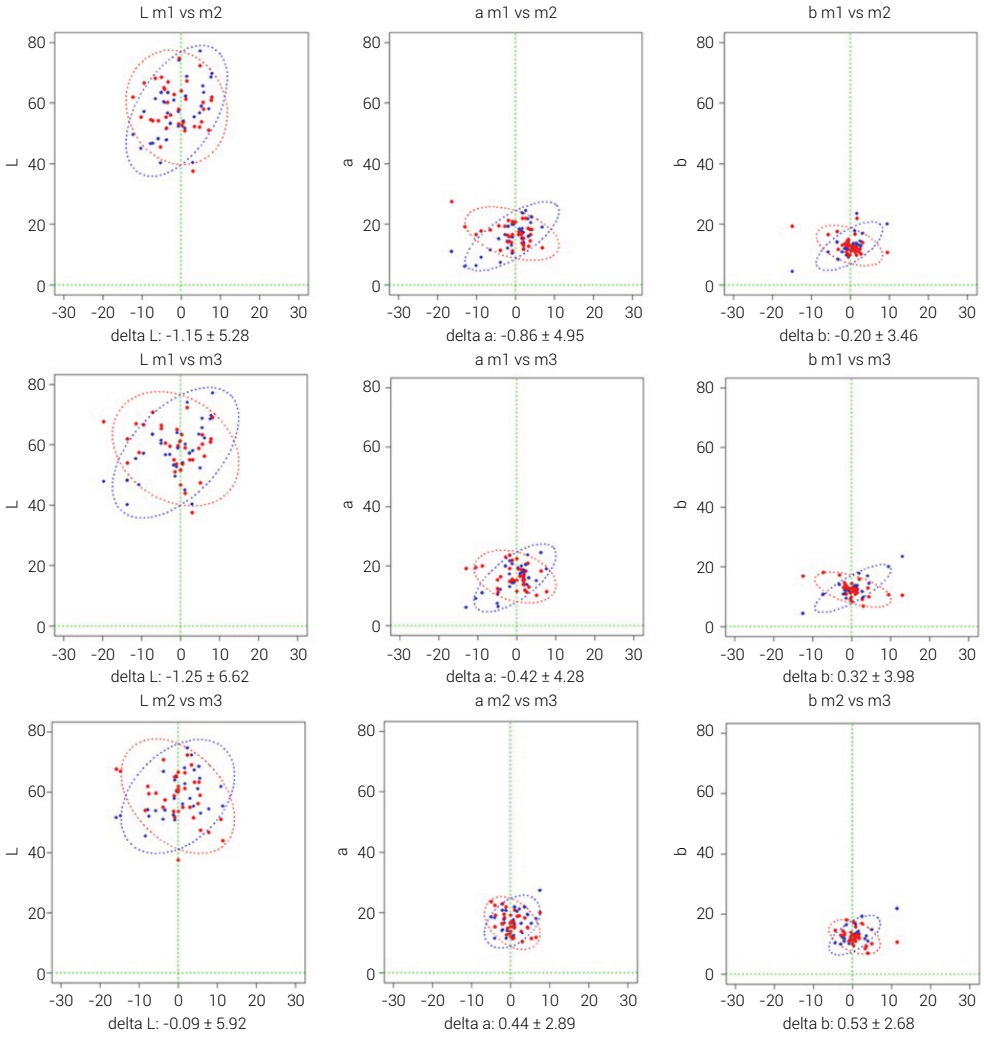

Figure 1. Comparison between measurements (to each measure blue or red) in the SPECTRO method (ellipses represent the 95% confidence interval) (R Project).

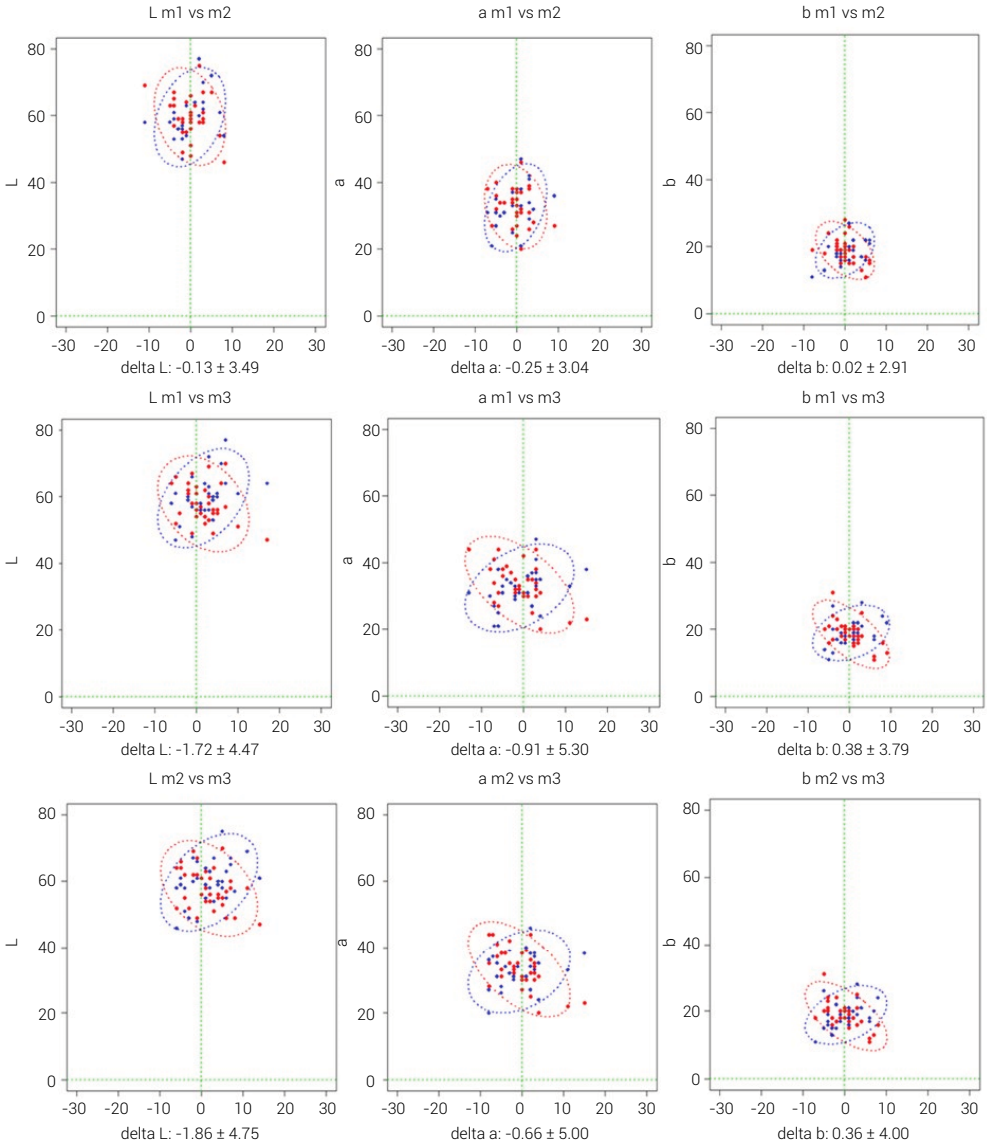

Figure 2. Comparison between measurements (to each measure blue or red) in the PHOTO method (ellipses represent the 95% confidence interval) (R Project).

Comparison between methods

The secondary hypothesis of the study is the comparison between methods. In Table 3 it was shown the Δs values for each method. The general mean was Δ: 20.5, higher than the expected value for the concordant methods. In the space color axes, the concordance between methods has presented that a* values do not acquire similarity in all measures (Table 4). The L* values had good concordance and only in one comparison, the b* value did not have a concordance (Table 4). Moreover, a* and b* values presented central and adequate distribution in most comparisons. One measure (m3) has shown a proportion bias (allocation trend outside the reference values), as was shown in Table 4. The comparison between methods can be viewed in the scatter diagram (Figure 3). It was evaluated the concordance coefficient, the value of the outlier, and the point intersections, as the greater the intersection of blue and red points, the greater the similarity between methods. The major intersection was presented by L* values, followed by b*.

Table 3. ΔE values for the comparison between methods PHOTO and SPECTRO.

ΔE between methods (n: 40)	M1	M2	M3
Mean ± sd (max.; min.)	20.4 ± 7.4 (33.2; 6.0)	19.8 ± 5.8 (34.1; 8.3)	21.38 ± 7.3 (35.1; 9.2)
Median, 25%,75%	21.7; 14. 7; 26.5	19.7; 15.8; 24.1	20.8; 15. 9; 27.9

max.: maximum; min.: minimum

Table 4. Bland-Altman concordance analysis between methods.

Spectro vs Photo n: 40	Pearson (r) correlation	p-value (paired t test/ one sample)	Mean of differences (sd)	ICC	Bland-Altman concordance (inferior; superior limits)	Proportion bias (linear simple regression) p-value for means
m1						
L	0.12	0.21	-2.13 (10.0)	0.11	-21.9 – 17.6	0.31
a	0.11	**<0. 05***	-16.3 (6.72)	0.02	-29.5 – -3.19	0.39
b	> 0.050#	1.00#	-6.11 (4.54)	0.02	-15.0 – 2.78	0.85
m2						
L	-0.04	0.50	-1.19 (9.92)	-0.04	-20.5 – 18.3	0.14
a	0.17	**<0. 05***	-15.7 (5.75)	0.02	-27.0 – -4.47	0.38
b	> 0.050#	**<0. 05#***	-5.88 (3.95)	0.08	-13.6 – 1.86	0.12
m3						
L	-0.31	0.65	0.83 (11.0)	-0.29	-20.8 – 22.5	0.94
a	0.11	**<0. 05***	-16.8 (6.5)	0.01	-29.7 – -3.95	**<0. 05§**
b	**0. 05#**	> 0.050#	-6.05 (5.2)	-0.25	-16.4 – 4.28	**<0. 05§**

Methods are not concordant; # b axis values show nonparametric distribution (Shapiro-Wilk test. p<0. 05): executed Wilcoxon test and Spearman correlation; ICC: intraclass correlation coefficient. ICC < 0.4 poor; 0.4 <= ICC < 0.75 satisfactory; ICC >= 0.75 excellent (Fleiss, J.L. The Design and analysis of clinical experiments. New York: Wiley, 1986). Linear regression significance of mean: if value was smaller than 0.05 there is a proportion bias§; (the reference levels had the tendency to concentrate above or below the averages / central reference), it means that the method tends to error only for high or low values.

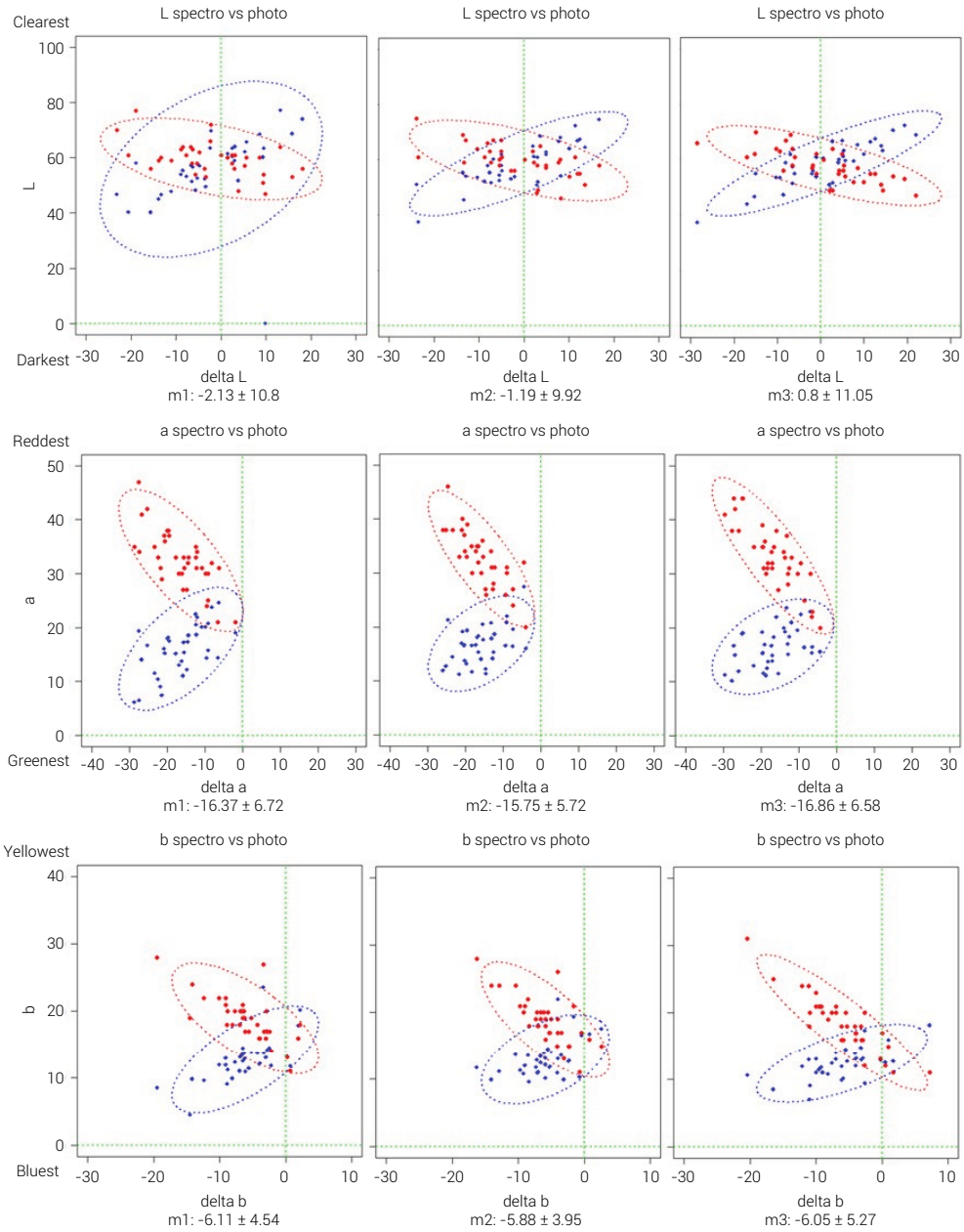

Figure 3. Scatter diagram, the SPECTRO method are represented by blue color and PHOTO method are represented by red color, (ellipses represent the 95% confidence interval) (R Project).

Focusing on the shape of the ellipses, the greater deformation of the ellipse, the greater the correlation between the X and Y coordinate values. Perfect circles indicated perfect independence and normality between measurement errors. It was possible to observe that the agreement index between methods was poor (Figure 3). However, in the intra method comparison, the measure distribution, confidence intervals, and the number of values of the outliers can be suitable for the not yet validated methods for gingival color evaluation. Also, when comparing ΔL*, Δa*,

and Δb* between methods (figure 3), it was observed that the biggest variation was in the a* values, approximately 16 units of difference compared to L* that had 2 and b* 5 units of difference.

Discussion

The level of agreement between the three measures of gingival color was evaluated and has exhibited better reproducibility and agreement grades in L*, a*, and b* axis, with adequate proportion and within the confidence intervals. For the ΔE values of 6.5 (SPECTRO) and 5.9 (PHOTO), the results assumed values above the threshold of the human eye (3.7) and those are already known for dental enamel, approximately ΔE: 3.3 (spectrophotometer) and ΔE: 2.9 (photography)[21,22]. Despite that, the quantification of gingival color by the tested methods has obtained an agreement level between the measures, and represents the main result of the study, mainly for L*, a*, and b*, when evaluated separately. It is emphasized that the methods were developed using conventional devices in the dental office (reflectance spectrophotometer to select enamel color and digital camera with software). The clinical relevance of this study was the research with common tools used in dentistry, the spectrophotometer (easyshade), and intra-oral photography to measure gingival color. One of the most used features recently in dentistry, photography can contribute to the auto evaluation of the state of health/disease of the patient, and as the gingival color can be measured through them and daily monitored by dentists after a periodontal plastic procedure to evaluate the cicatricial tissue course, inflammation levels, and aesthetic.

The devices used on the tested methods are alternatives[36], when compared to studies that use colorimeters[10] or spectrophotometers[12,37] , specific for gingival scanning and entail additional costs. This explains how the ΔE values and the sensibility of both methods have exceeded the threshold of the human eye. On the other hand, the quantification of gingival color was possible, having a great potential for future use in research. In addition, usual and common tools in the dental office were also used and have presented with an acceptable agreement level between the measures. Spectrophotometers has already been used in dental rehabilitation and dental bleaching to evaluate dental color. The present study has shown a possible use of this tool in periodontology and implant dentistry, for the evaluation of gingival color, having acceptable agreement and concordance rates. In the PHOTO method, a digital camera or cell phone camera is able to acquire the images, but in order to reproduce the results observed in the study, it is necessary to have access to a payable software, in order to execute the examinations of L* a* b* values. Despite that, mobile apps with color scanning functions are free and available on different digital platforms. Even though these tools have not been yet validated for gingival color measurement, they may be tested and therefore, expand the universalization of the method.

In the first hypothesis (intra method comparison), in the systematic error (evaluate the method accuracy measured by presence or absence of bias) and in the random error (accuracy between measurements)[38,39], the bias or systematic error was evaluated by continuous values. In the Wilcoxon or on the T- tests, it was revealed that in

the PHOTO group there was bias between the L* axis values (p<0,05), differently to the SPECTRO group, which the axis needs to be adjusted to maintain similarly in the photo acquisition protocol, which interferes in the reading the other outcomes a * and b *. The difficulty of controlling the luminosity can explain the observed difference in the L* axis, beyond the level of sensibility of software to capture minimal different values of L*.

A random error is not predictable and it uses the estimate through the agreement coefficients (Bland-Altman). Measures differences were not observed in this concept, since all the limits of agreement were <1.7, and the agreement coefficients (value close to the maximum agreement) presented similar intervals between the measures (maximum of 25 and minimum of 12 units). The b* axis represented the minor interval of agreement coefficients and the lowest limit of agreement, thus obtaining the major approximation to the perfect concordance among measures. The a* axis presented intermediate agreements intervals and the L * axis presented the largest agreement intervals between the measures and the highest agreement limits, values that deviate from the perfect agreement (perfect agreement order: b*>a*>L*).

Regarding what was observed in the systematic error, the SPECTRO method had major reliability in different measures. For a method to be reliable, the systematic and random errors must be known and contained in the statistical limit of difference. Using the concordance limits has the benefit of not requiring data with parametric distribution and fewer comparisons[27]. To measure the "strength" of reliability, Pearson or Spearman's correlation coefficient was used and the similarity was considered satisfactory between measurements (ICC mean: 0.44 SPECTRO/0.57 PHOTO). However, this analysis had limitations, since only the values were used to measure the agreement between methods[32]. The real interpretation of this concept is that the differences between the measurements were not large enough to be detected in the sample size. Thus, the model that best express all the information about the comparisons is based on two-dimensional scatter plots with confidence intervals[33].

The second study hypothesis analyzes the sensibility between the SPECTRO and the PHOTO methods. For the a* values, the results have not shown a good agreement among methods. However, for the L* and in most of b* values, a reasonable agreement was observed. For the dental enamel evaluation, the a* values have varied beyond expectations when compared in the same methods and the values of L* and b* have shown an excellent level of agreement[8]. The results favor the photography method (plus software) as an alternative, compared to the spectrophotometer, for the reliable acquisition of the variables L* and b*.

The agreement level of any measurement method needs information. In periodontology, the probing depth exam helps in the clinical identification of periodontal parameters (sulcus or pocket probing depth, clinical attachment level, beyond bleeding upon probing, and inflammation signals). Thus, agreement and sensitivity of the method of measurement combined with the instruments help the operator in identifying the disease/health outcomes. In comparison, the level of error obtained between probing exams is above 1 mm for systematic error and between 0.3 to 0.7 mm as a random error for establishing the attachment clinical level. Among evaluators with both manual and electronic probes, the intra-class correlation coef-

ficient is from 0.41 to 0.90 (reasonable to excellent)[8]. Even a conventional instrument shows changes in its measurements depending on the operators. Determining all types of instruments variation is essential to establish fair regimes and the most adequate research protocols.

The use of photographs is not recent in dentistry[40], not even in the study of tooth enamel color[6]. New tools were included to facilitate the outcomes from the collection of disorders and diseases[41]. Intraoral scanners are the most current technology for dental impressions and on acquiring oral characteristics. Depending on the file format generated, colors are also present in this analysis. Nevertheless, it is not yet possible to use the polychromatic scanned files for color evaluation with quality in analysis[42,43]. Despite that, with advances and improvements in technology, it is not difficult to imagine that color will be another factor better incorporated into these tools. When this alternative is available in a quality, the validation for color analysis will also be necessary, even for the comparison between methods and their sensitivity. Thus, the next steps for understanding and validating methods used in this study (and their execution format) are the comparison between different operators of the software and the photographs. Mainly due to the PHOTO method that had more variables and still needs to be continually tested for its effective potential and for being the cheapest gingival color analysis tool[21]. Even though the used spectrophotometer had a lower number of variables also needs calibrations focused on colors of the natural gingiva, with comparison to directories related to the race, age, and sex/gender of patients[13,14]. With these elements, the accuracy of the methods would be effectively tested, and the quality of the results better debated.

This study had limitations, such as the photograph protocol and adjustments beyond the patient head position[7], that could interfere in the measurement[21,22]. The SPECTRO method needed a better calibration system, aiming to measure gingival color. Another limitation was in the point of both analyses of the methods, which was executed 2 mm the gingival margin. This point was defined without guides and the periodontal phenotype was not evaluated as a possible interference. Also, comparison with specific colorimeters[10] or spectrophotometers[12,37], able to evaluate the gingival color, was not executed. Nevertheless, this study protocol could compare three different measures, using for each method a feasible comparison pair and an adjusted and complementary statistical analysis system described by previous studies[26,27,29-33,38,39].

In conclusion, both methods could quantify the gingival color from the coordinates L *, a *, and b *. The evaluation of the intra method has shown slight variations between the measurements and greater reliability for the SPECTRO method. The comparison between methods showed little agreement between them, mainly for a* values.

Acknowledgments

We are grateful to Professor Dr. Adilson Yoshio Furuse (furuse@usp.br) from the Dentistry discipline of the Department of Dentistry, Endodontics and Dental Materials at the Bauru School of Dentistry -the University of São Paulo, to loan borrowing the spectrophotometer used throughout our study. Also, Dr. Shin-Jae Lee (nonext.shinjae@gmail.com) and Dr. Richard Donatelli (RDONATELLI@dental.ufl.edu) for the prompt help and availability of the R language scripts.

Data availability

Datasets related to this article will be available upon request to the corresponding author

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 (process number 88882.182735/2018-1), and National Council for Scientific and Technological Development – CNPq/PIBIC (N° 653-process number119839/2019-2).

Disclosure statement

The authors declare that there is no conflict of interest.

Author Contribution

M.V.C. and C.A.D. designed the study model, the computational framework, and analyzed the data. M.V.C. and G.V. carried out the implementation. M.S.R.Z. performed the calculations. M.V.C. and A.C.P.S. wrote the manuscript with input from all authors. M.V.C. and G.V. conceived the study and were in charge of overall direction and planning.

All authors actively participated in the discussion of the manuscript's findings and have revised and approved the final version of the manuscript.

References

1. Bach Junior J. [Physiological colors in the science of Goethe: education and phenomenology]. Cien Educ (Bauru). 2016 Mar;22(1):117-28. Portuguese. doi: 10.1590/1516-731320160010008.
2. Brewer JD, Wee A, Seghi R. Advances in color matching. Dent Clin North Am. 2004 Apr;48(2):v, 341-58. doi: 10.1016/j.cden.2004.01.004.
3. Heydecke G, Schnitzer S, Türp JC. The color of human gingiva and mucosa: visual measurement and description of distribution. Clin Oral Investig. 2005 Dec;9(4):257-65. doi: 10.1007/s00784-005-0006-3.
4. Schnitzer S, Türp JC, Heydecke G. Color distribution and visual color assessment of human gingiva and mucosa: a systematic review of the literature. Int J Prosthodont. 2004 May-Jun;17(3):327-32.
5. Johnston WM, Kao EC. Assessment of appearance match by visual observation and clinical colorimetry. J Dent Res. 1989;68(5):819-22. doi: 10.1177/00220345890680051301.
6. Chu SJ, Trushkowsky RD, Paravina RD. Dental color matching instruments and systems. Review of clinical and research aspects. J Dent. 2010;38 Suppl 2:e2-e16. doi: 10.1016/j.jdent.2010.07.001.
7. Wee AG, Lindsey DT, Kuo S, Johnston WM. Color accuracy of commercial digital cameras for use in dentistry. Dent Mater. 2006 Jun;22(6):553-9. doi: 10.1016/j.dental.2005.05.011.
8. Anand D, Kumar GP, Anand DY, Sundar MK, Sharma R, Gaurav A. Shade selection: Spectrophotometer vs. digital camera–A comparative in vitro study. Ann Prosthod Restor Dent. 2016 Jul;2(3):73-8.
9. Hein S, Tapia J, Bazos P. eLABor_aid: a new approach to digital shade management. Int J Esthet Dent. 2017;12(2):186-202.

10. Hyun HK, Kim S, Lee C, Shin TJ, Kim YJ. Colorimetric distribution of human attached gingiva and alveolar mucosa. J Prosthet Dent. 2017;117(2):294-302. doi: 10.1016/j.prosdent.2016.06.009.

11. Ho DK, Ghinea R, Herrera LJ, Angelov N, Paravina RD. Color range and color distribution of healthy human gingiva: a prospective clinical study. Sci Rep. 2015 Dec 22;5:18498. doi: 10.1038/srep18498.

12. Huang JW, Chen WC, Huang TK, Fu PS, Lai PL, Tsai CF, Hung CC. Using a spectrophotometric study of human gingival colour distribution to develop a shade guide. J Dent. 2011 Dec;39 Suppl 3:e11-6. doi: 10.1016/j.jdent.2011.10.001.

13. Ito M, Marx DB, Cheng AC, Wee AG. Proposed shade guide for attached gingiva--a pilot study. J Prosthodont. 2015;24(3):182-7. doi:10.1111/jopr.12195.

14. Burak Ak, Kaki B. Color distribution of gingiva created by color schema maker software. East J Med. 2016 Apr;21(2):69-74. doi: I:10.5505/ejm.2016.32032.

15. Cortellini P, Bissada NF. Mucogingival conditions in the natural dentition: Narrative review, case definitions, and diagnostic considerations. J Periodontol. 2018;89 Suppl 1:S204-13. doi: 10.1002/JPER.16-0671.

16. Jung RE, Siegenthaler DW, Hämmerle CH. Postextraction tissue management: a soft tissue punch technique. Int J Periodontics Restorative Dent. 2004 Dec;24(6):545-53.

17. Jung RE, Holderegger C, Sailer I, Khraisat A, Suter A, Hämmerle CH. The effect of all-ceramic and porcelain-fused-to-metal restorations on marginal peri-implant soft tissue color: a randomized controlled clinical trial. Int J Periodontics Restorative Dent. 2008;28(4):357-65.

18. Bayindir F, Bayindir YZ, Gozalo-Diaz DJ, Wee AG. Coverage error of gingival shade guide systems in measuring color of attached anterior gingiva. J Prosthet Dent. 2009 Jan;101(1):46-53. doi: 10.1016/s0022-3913(08)60290-5.

19. Mayer Y, Ginesin O, Machtei EE. Photometric CIELAB analysis of the gingiva: a novel approach to assess response to periodontal therapy. J Periodontol. 2017;88(9):854-9. doi: 10.1902/jop.2017.170004.

20. Sailer I, Fehmer V, Ioannidis A, Hämmerle CH, Thoma DS. Threshold value for the perception of color changes of human gingiva. Int J Periodontics Restorative Dent. 2014;34(6):757-62. doi: 10.11607/prd.2174.

21. Knösel M, Attin R, Jung K, Brunner E, Kubein-Meesenburg D, Attin T. Digital image color analysis compared to direct CIE colorimeter assessment under different ambient conditions. Am J Dent. 2009 Apr;22(2):67-72.

22. Sluzker A, Knoesel M, Athanasiou AE. Sensitivity of digital dental photo CIE L⋆ a⋆ b⋆ analysis compared to spectrophotometer clinical assessments over 6 months. Am J Dent. 2011 Oct 1;24(5):300-4.

23. Cairo F, Nieri M, Cattabriga M, Cortellini P, De Paoli S, De Sanctis M, et al. Root coverage esthetic score after treatment of gingival recession: an interrater agreement multicenter study. J Periodontol. 2010 Dec;81(12):1752-8. doi: 10.1902/jop.2010.100278.

24. Luo MR. A colour management framework for medical imaging applications. Comput Med Imaging and Graph. 2006;30(6-7):357-61. doi: 10.1016/j.compmedimag.2006.09.002.

25. Judd DB, Wyszecki G. Color in business, science and industry. 3 ed. New York: John Wiley; 1975 p.332.

26. Springate SD. The effect of sample size and bias on the reliability of estimates of error: a comparative study of Dahlberg's formula. Eur J Orthod. 2012;34(2):158-63. doi: 10.1093/ejo/cjr010.

27. Martelli Filho JA, Maltagliati LÁ, Trevisan F, Gil CT. [New statistical methods to evaluate reproducibility]. Rev Dental Press Ortod Ortop Facial. 2005 Oct;10(5):122-9. doi: 10.1590/S1415-54192005000500012.

28. Houston WJ. The analysis of errors in orthodontic measurements. Am J Orthod. 1983 May;83(5):382-90. doi: 10.1016/0002-9416(83)90322-6.

29. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989 Mar;45(1):255-68.

30. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135-60. doi: 10.1177/096228029900800204.

31. Dahlberg G. Statistical Methods for Medical and Biological Students. Br Med J. 1940 Sep;2(4158):358-9.

32. Donatelli RE, Lee SJ. How to report reliability in orthodontic research: Part 1. Am J Orthod Dentofacial Orthop. 2013;144(1):156-61. doi: 10.1016/j.ajodo.2013.03.014.

33. Donatelli RE, Lee SJ. How to report reliability in orthodontic research: Part 2. Am J Orthod Dentofacial Orthop. 2013;144(2):315-8. doi: 10.1016/j.ajodo.2013.03.023.

34. IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.; 2019.

35. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013 [cited 2021 Nov 12]. Available from: http://www.R-project.org/.

36. Mehl A, Bosch G, Fischer C, Ender A. In vivo tooth-color measurement with a new 3D intraoral scanning system in comparison to conventional digital and visual color determination methods. Int J Comput Dent. 2017;20(4):343-61.

37. Gil MS, Ishikawa-Nagai S, Elani HW, Da Silva JD, Kim DM, Tarnow D, et al. A prospective clinical trial to assess the optical efficacy of pink neck implants and pink abutments on soft tissue esthetics. J Esthet Restor Dent. 2017;29(6):409-15. doi: 10.1111/jerd.12309.

38. Perinetti G. StaTips Part I: Choosing statistical test when dealing with differences. South Eur J Orthod Dentofac Res. 2016 Apr;3(1):3-4. doi: 10.5937/sejodr3-1264.

39. Perinetti G. StaTips Part II: Assessment of the repeatability of measurements for continuous data. South Eur J Orthod Dentofac Res. 2016 Oct;3(2):33-4.

40. Kerner S, Etienne D, Malet J, Mora F, Monnet-Corti V, Bouchard P. Root coverage assessment: validity and reproducibility of an image analysis system. J Clin Periodontol. 2007;34:969-76. doi: 10.1111/j.1600-051X.2007.01137. x.

41. Badano A, Revie C, Casertano A, Cheng WC, Green P, Kimpe T, et al. Consistency and standardization of color in medical imaging: a consensus report. J Digit Imaging. 2015;28(1):41-52. doi: 10.1007/s10278-014-9721-0.

42. Gotfredsen K, Gram M, Ben Brahem E, Hosseini M, Petkov M, Sitorovic M. Effectiveness of shade measurements using a scanning and computer software system: a pilot study. Int J Oral Dent Health. 2015;1(008). doi: 10.23937/2469-5734/1510008.

43. Yoon HI, Bae JW, Park JM, Chun YS, Kim MA, Kim M. A study on possibility of clinical application for color measurements of shade guides using an intraoral digital scanner. J Prosthodont. 2018;27(7):670-5. doi: 10.1111/jopr.12559.