ARTICLE IN PRESS

Dental Materials xxx (xxxx) xxx

Contents lists available at ScienceDirect

Dental Materials

journal homepage: www.elsevier.com/locate/dental

Bridging instrumental and visual perception with improved color difference equations: A multi-center study

Sascha Hein [a,*], Omnia Saleh [b,d], Changjun Li [c], Julian Nold [d], Stephen Westland [a]

[a] School of Design, University of Leeds, Leeds, UK
[b] Prosthodontic Division, Department for Restorative Science and Biomaterials, Boston University Henry M. Goldman School of Dental Medicine, Boston, USA
[c] School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan, China
[d] Medical Center, University of Freiburg, Center for Dental Medicine, Department of Prosthetic Dentistry, Faculty of Medicine, University of Freiburg, Freiburg, Germany

A R T I C L E I N F O

Keywords:
Dentistry
Shade matching
Tooth color measurement
Optimized color difference equations
Visual-instrumental agreement
Visual perception
STRESS index

A B S T R A C T

Objectives: This multicenter study aimed to evaluate visual-instrumental agreement of six color measurement devices and optimize three color difference equations using a dataset of visual color differences (ΔV) from expert observers.

Methods: A total of 154 expert observers from 16 sites across 5 countries participated, providing visual scaling on 26 sample pairs of artificial teeth using magnitude estimation. Three color difference equations (ΔE^*_{ab}, ΔE_{00}, and CAM16-UCS) were tested. Optimization of all three equations was performed using device-specific weights, and the standardized residual sum of squares (*STRESS*) index was used to evaluate visual-instrumental agreement.

Results: The ΔE^*_{ab} formula exhibited *STRESS* values from 18 to 40, with visual-instrumental agreement between 60 % and 82 %. The ΔE_{00} formula showed *STRESS* values from 26 to 32, representing visual-instrumental agreement of 68 % to 74 %. CAM16-UCS demonstrated *STRESS* values from 32 – 39, with visual-instrumental agreement between 61–68 %. Following optimization, *STRESS* values decreased for all three formulas, with ΔE^{*} demonstrating average visual-instrumental agreement of 79 % and ΔE_{00} of 78 %. CAM16-UCS showed average visual-instrumental agreement of 76 % post optimization.

Significance: Optimization of color difference equations notably improved visual-instrumental agreement, overshadowing device performance. The optimzed ΔE^{*} formula demonstrated the best overall performance combining computational simplcity with outstanding visual-instrumental agreement.

1. Introduction

Over recent decades, instrumental color measurement in dentistry has gained prominence for its objectivity and precision in assessing tooth color [1,2]. However, the exclusive focus on instrumental measurements often neglects the critical role of visual color perception. A notable challenge arises when there is a discrepancy between instrumental measurements and visual perception, leading to situations where a restoration appears visually dissimilar despite a small, measured color difference. Typically, the accuracy of tooth color measurements is not only dependent on the device's performance but also significantly influenced by the color difference equations used. This raises critical questions: are discrepancies between visual and instrumental evaluations the result of the device's capabilities, or the equations applied? And how do we disentangle these two things?

The widespread use of CIELAB color space in dentistry [3–5] has spurred extensive research comparing color measurement instruments to establish their relative accuracy [6,7]. Yet, within dental colorimetry research, confusion persists between the concepts of accuracy and precision, sparking debates over the most reliable instruments [8–11]. Lack of precision, stemming from random noise, differs from lack of accuracy, which results from systematic bias [12]. True accuracy in color measurement requires calibration against recognized standards. These standards are usually measured by a national standardizing laboratory equipped with the finest instrumentation and procedures. Each standard comes with a certificate detailing an estimate of the associated measurement uncertainty, providing a definitive benchmark for evaluating the accuracy of color measurements [13,14]. In tooth color measurement, accuracy is usually defined as the system's ability to record the 'true' CIELAB values, yet determining these values poses challenges [5].

* Correspondence to: School of Design, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK.
E-mail address: sdsch@leeds.ac.uk (S. Hein).

https://doi.org/10.1016/j.dental.2024.07.003
Received 18 April 2024; Accepted 15 July 2024

The perennial quest to describe visual color differences with a single distance metric began with the conception of CIELAB in 1976 as an approximately uniform color space, within which equal visual color differences are represented by equal Euclidean distances (ΔE^*_{ab}) [15]. However, it soon became evident that ΔE^*_{ab} fell short of expectations [16], catalyzing the development of more advanced color difference equations such as the CIEDE2000 (ΔE_{00}) [17] which is currently the CIE recommendation for small color differences [18]. However, such efforts have largely focused on color match specifications rather than perceptual color appearance properties [19]. This has led to the development of more advanced colorimetry such as CAM16-UCS [20] which has been the subject of recent investigations in dental research [21,22].

Given the evolution of color difference equations, several methods are now available for instrumental color difference evaluation. These range from straightforward computations to those considered state-of-the-art technological advancements, featuring considerable complexity. However, it remains unclear which of these methods offer the best visual-instrumental agreement, posing an important question for dental practitioners seeking reliable color measurement solutions for clinical applications.

Therefore, the aim of this study is to investigate visual-instrumental agreement of six devices and to optimize three color difference equations, using a large visual dataset from expert observers collected as a multi-center study. The study explores two distinct null hypotheses to thoroughly assess the performance and optimization of color difference equations in capturing perceived color differences through instrumental evaluation:

1. There is no significant difference in visual-instrumental agreement among the tested devices in reflecting perceived color differences.
2. Optimized color difference equations offer no significant improvement in performance over their generic counterparts.

By clarifying which devices and equations may offer improved visual-instrumental agreement, this study aims to support dental professionals in making informed decisions regarding color measurement in clinical practice, ultimately contributing to enhanced patient care and treatment outcomes.

2. Material and methods

2.1. Visual scaling technique

To examine visual color differences (ΔV), a suitable scaling technique needs to be applied to quantify visual differences between pairs of teeth. The gathered ΔV data plays a key role in assessing the performance of color measurement devices by examining the correlation between instrumental color difference (ΔE) and ΔV [23]. A technique commonly referred to as magnitude estimation (ME) was selected as the visual scaling technique for this study due to its successful application in color-difference research [24]. This technique involves expert observers assigning numerical values to perceived color differences, thus enabling a consistent and quantifiable assessment of ΔV [25]. The robustness and reliability of ME in quantifying subtle color differences are particularly relevant in dentistry. This is because the color differences typically encountered in this field are relatively small [26], in contrast to the larger color differences often found in visual datasets from other industries [27].

2.2. Selection of expert observers

In preparation for this study, an application for proportional ethical review was submitted and subsequently granted (Ethical approval number LTDESN-196). To fulfill the specialized needs for ME, 154 expert observers were recruited who passed the Ishihara test for color deficiency. They consisted of dental practitioners and dental technicians with relevant experience in the field of restorative dentistry. The demographic overview (Table 1) presents data on 105 dental practitioners and 49 dental technicians.

2.3. Selection of multi-center sites

Sites were strategically chosen to be reflective of a broad spectrum of professional settings. The 16 centers participating in this investigation include renowned universities with dedicated dental schools, specialized private dental laboratories, and dental practices known for their excellence in dental care and research. Table 2 provides an overview of the locations and types of centers that contributed to this study.

2.4. Visually scaled samples

Four hyper-realistic phantom models were custom fabricated by an experienced master dental technician. These models were made of micro filler reinforced composite denture teeth (Physiodens, Vita Zahnfarbrik, Germany) as depicted in Fig. 1. To analyze the correlation between visually perceived color differences and those calculated from each device, colorimetric data from appropriate sample pairs was necessary. Four color centers within CIELAB color space were identified for this study, as shown in Fig. 2. Each phantom model, representing one base shade (1M2, 2M2, 3M2, and 4M2), was combined with 5 to 7 exchangeable teeth per model, to create 26 visually scaled sample pairs. This regime led to the creation of four color-centers of different lightness, where sample pairs primarily differed in hue and chroma, accommodating findings that simultaneous assessment of lightness and chromaticness can increase observational uncertainty [28]. The color difference between each of the 26 sample pairs was less than 5 ΔE^*_{ab} units [18].

2.5. Psychophysical experiment

To determine visual-instrumental agreement, a psychophysical experiment was carried out under controlled conditions. Observers viewed sample pairs at a distance of approximately 35–50 cm against a 45º angled surface painted in Munsell N5 neutral grey (GTI Gmbh, Harrislee, Germany). Simulated daylight of 6500 K was provided by a viewing cabinet (DLS Color Viewing Light v7, JustNormlicht, Germany) with an illuminance of approximately 1000 lx [29]. The setup allowed observations while standing in a darkened room. In accordance with the ME technique, each participant was then asked to rate the sample match for each pair from 0 % (worst match) to 100 % (perfect match). Phantom models were used in random order, with each maxillary central tooth drawn from a bag without replacement. Responses were recorded in an Excel sheet (Microsoft Corp., Redmond, WA, USA) using custom drop-down menus for consistency in data collection. Each participant's session lasted approximately 20–30 min.

2.6. Instrumental measurement of sample pairs

In this study, a range of color measurement devices frequently cited in dental literature for tooth color assessment were evaluated and they are listed in Table 3. The selection included both established systems and newer technologies for which there is currently limited data available. Sample pairs were mostly measured with devices specifically designed for dentistry, each with its own illumination geometry and straightforward measurement regime.

Exceptions to this standard procedure involved two systems:

1. **Tele-radiospectrometer (PR-670):** Measurements were taken using a calibrated tele spectroradiometer (SpectraScan PR-670, Photo Research Inc., Syracuse, NY, USA) with a 1º aperture and a 45º:0º illumination geometry provided by the same viewing cabinet that was used for the visual experiment.

Table 1
Gender and age (yrs) distribution of expert observers participating in this study.

Profession	n	Male	Female	18-24 yrs	25-34 yrs	35-44 yrs	45-54 yrs	55-64 yrs	≥ 65 yrs
Dental practitioners	105	56	49	6	64	12	11	8	0
Dental technicians	49	29	20	2	12	22	11	5	1

Table 2
Participating centers in the multi-center study, including location and type of institution.

Country	City/ Region	Center Name	Type
Austria	Oetztal	Die Zahnmanufaktur	Private Dental Laboratory
France	Strasbourg	University of Strasbourg	Faculty of Dental Medicine
Germany	Cham	Cham Zahntechnik	Private Dental Laboratory
Germany	Munich	Ludwig-Maximilians-University Munich	School of Dental Medicine
Germany	Landshut	Hofmann Dentaltechnik GmbH	Private Dental Laboratory
Germany	Erlstaett	Oral Design Chiemsee GmbH	Private Dental Laboratory
Germany	Freiburg	Albert-Ludwigs-University Freiburg	School of Dental Medicine
Germany	Frankfurt	Goethe University	School of Dental Medicine
Germany	Bonn	University of Bonn	School of Dental Medicine
Germany	Duesseldorf	Heinrich Heine University	School of Dental Medicine
Switzerland	Bern	Zahnmanufaktur Zimmermann & Maeder AG	Private Dental Laboratory
Switzerland	Bern	Praxis Mathey	Private Dental Practice
Switzerland	Bern	University of Bern	School of Dental Medicine
Switzerland	Zurich	University of Zurich	School of Dental Medicine
United Kingdom	London	University College London	Eastman Dental Institute
United Kingdom	Leeds	University of Leeds	NHS Teaching Hospital

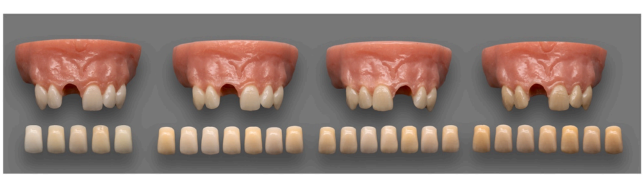

Fig. 1. Hyper-realistic phantom models were fabricated each in one base shade (from left to right: 1M2, 2M2, 3M2, and 4M2) to closely mimic natural teeth and for facilitating a more realistic assessment of color differences in a clinical setting.

2. **eLAB System (eLAB):** Not a device but a system for calibrating digital RAW images. Samples were photographed using a Nikon D7500 digital camera equipped with a 105 mm macro lens (Nikon Corp., Germany) and a ring flash (MK-14EXT, Meike, Germany) alongside a cross-polarization filter (polar_eyes, Emulation, Germany). The samples were positioned within the viewing cabinet, with a grey card (white_balance, Emulation, Germany) placed beneath the incisal edges to ensure consistent exposure settings. The eLAB protocol was stringently followed, setting the aperture at f22, exposure time at 1/125 s, ISO at 100, and using the RAW image format [30]. Subsequent processing and calibration in the eLAB software allowed for CIELAB value measurements.

To account for variability across the tooth surface, each sample was measured three times in each of three distinct regions and averaged: cervical, middle, and incisal areas of the labial tooth surface.

2.7. Computation of color differences between sample pairs

In the scope of this research, distinguishing between the performance of color measurement devices and the efficacy of color difference equations poses a significant challenge. To ensure consistency across all assessments, all color differences were computed under Illuminant D65 and for the CIE 1931 standard colorimetric observer [31]. Three color difference metrics were then employed as baseline assessments:

1. **ΔE^*_{ab}:** Utilizes the Euclidean distance for color difference calculation, providing a straightforward approach to quantifying color variations.
2. **ΔE_{00}:** Incorporates weighting functions S_L, S_C, S_H, each set to 1, enhancing the model's sensitivity to hue, chroma, and lightness differences.
3. **CAM16-UCS:** A uniform color space that designates J for lightness and a and b for chromaticity coordinates, indicating redness-greenness and yellowness-blueness, respectively. Since visual observations were carried out in a viewing cabinet, luminance levels for each instrument were considered to be the same, defining the surround parameters as 'average' with $F = 1.0$, $c = 0.69$ and $N_c = 1$. The background parameter was set to $Y_B = 20$ due to the neutral grey paint against which the samples were viewed ($L^* = 50$). The luminance level provided by the viewing cabinet was approximately 1000 lx, therefore the reference white in the reference illuminant was set to $Y_W = 100$ with an adaptation luminance of $L_A = 64$ cd/m^2 [20].

2.8. Statistical analysis

Data evaluation was performed using a specialized color toolbox in MATLAB (R2023b; MathWorks, Natick, MA, USA), which provides various functions for computational color science [32]. A MATLAB routine was also used to access the Excel sheet containing anonymized participant data, enabling efficient extraction of relevant information. The visually scaled color differences were determined by calculating geometric and arithmetic means, as well as the median for each observer response. The performance of each color measurement device was assessed by computing color differences for each sample pair. The standardized residual sum of squares (*STRESS*) index was used for performance evaluation where values ranging from 0 to 100 are indicative of device performance, with lower values signaling better visual-instrumental agreement [33]. Conversely, 100 - *STRESS* provides a direct measure of visual-instrumental agreement [34]:

$$STRESS = 100\left(\frac{\sum_i(\Delta E_i - F_1\Delta V_i)^2}{\sum_i F_i^2\Delta V_i^2}\right)^{1/2} \text{ and } F = \frac{\sum_i \Delta E_i^2}{\sum_i \Delta E_i \Delta V_i}$$

STRESS can also be used to express observer variability or the difference in performance between two devices since the square of the ratio of *STRESS* values from two visual data sets follows a two-tailed *F*-distribution, as it is equivalent to the ratio of two chi-squared variables [35]. In simpler terms, when comparing the performance of two devices, the *STRESS* value is analyzed, which follows a specific statistical distribution.

In statistical terms, this distribution adheres to an *F*-variable, where a

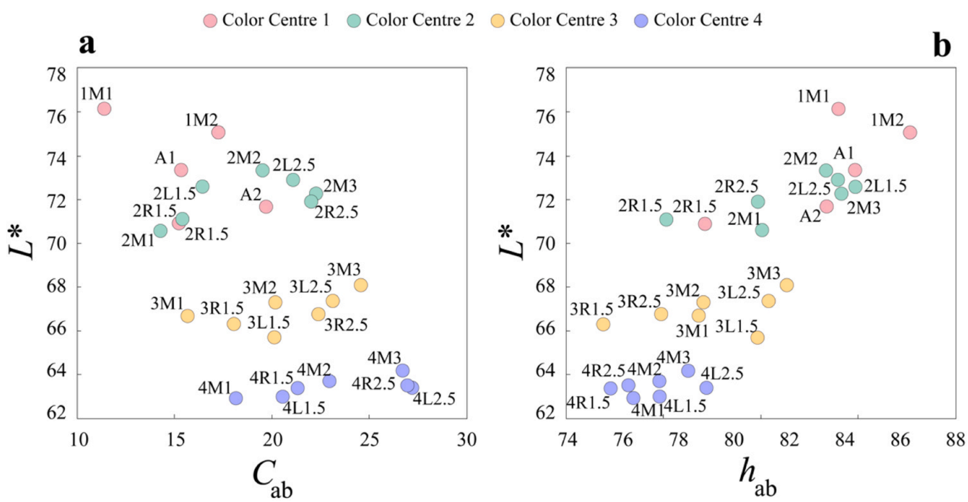

Fig. 2. Color coordinates for 26 visually scaled samples plotted by their lightness (L^*) and chroma (C^*_{ab}) distribution (A) and by their lightness (L^*) and hue angle (h_{ab}) distribution (B). Four color centers were identified, corresponding to the shades 1M2 (Color Center 1), 2M2 (Color Center 2), 3M2 (Color Center 3), and 4M2 (Color Center 4).

Table 3
Color measurement devices investigated in this study, including device name, manufacturer, geographical location of manufacturer, and type of device.

Device Name	Manufacturer	Location	Type
SpectroShade Micro II (SSM II)	Spectroshade USA	Oxnard, CA, USA	Multispectral camera
SpectraScan PR–670 (PR–670)	Photo Research Inc.	Syracuse, NY, USA	Tele radiospectrometer
Rayplicker Cobra (RPC)	Borea	France	Multispectral camera
Optishade (OS)	Smile Line	Switzerland	Calibrated camera
eLAB & Nikon D7500 (eLAB)	Emulation	Germany	Calibrated camera
Vita Easy Shade V (ES-V)	Vita Zahnfabrik	Germany	Photo spectrometer

critical value (F_C) denotes the threshold for rejecting the null hypothesis. The null hypothesis, in this context, suggests that two devices (A and B) exhibit no significant differences. To evaluate this hypothesis, the *F*-value is derived from the *STRESS* index:

$$F = \frac{STRESS_A^2}{STRESS_B^2}$$

If the *F*-value falls below a certain critical threshold ($F<F_C$) or exceeds the inverse of that threshold ($F>1/F_C$), the null hypothesis must be rejected. This critical threshold for F_C is determined by the two-tailed *F*-distribution with a 95 % confidence level and degrees of freedom (N − 1, N − 1), where N represents the sample size. In the present case this results in:

$$F_C = 1.955$$

$$\frac{1}{F_C} = 0.512$$

In addition to *STRESS*, the correlation between computed and visually scaled color differences was assessed using the R-squared method to further elucidate the relationship between instrumental measurements and visual assessments.

2.9. Optimization of color difference equations

It has been suggested that a correction of color difference equations can be used to improve their performance, in particular for very small color differences [34]. For this purpose, initial metrics, ΔE^*_{ab}, ΔE_{00} and CAM16-UCS, underwent optimization to enhance their alignment with perceptual color differences. First, the conventional ΔE^*_{ab} formula was subjected to a targeted optimization process:

$$\Delta E' = \sqrt{S_L(L_1 - L_2)^2 + S_C(a_1 - a_2)^2 + S_H(b_1 - b_2)^2}$$

where ΔE’ is the improved color distance equation to achieve better congruency between computed and visually perceived color difference by optimizing parameters S_L, S_C, S_H.

For ΔE_{00}, optimization was pursued through the tailored adjustment of its weighting functions: S_L, S_C, and S_H:

$$\Delta E_{00} = \left[\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T\left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)\right]^{1/2}$$

Similarly, CAM16-UCS underwent optimization, also with tailored parameters S_L, S_C, S_H:

$$\Delta \overline{E} = \sqrt{S_L(J_1 - J_2)^2 + S_C(a_1 - a_2)^2 + S_H(b_1 - b_2)^2}$$

These adjustments are device-specific, acknowledging the unique color measurement capabilities and limitations inherent to each device. In all three cases the *fminsearch* function from the MATLAB optimization toolbox was used to find the optimal parameters as evaluated by the *STRESS* index.

3. Results

3.1. Visual-instrumental agreement

Visually scaled color differences from 154 observers for 26 sample pairs were averaged using the arithmetic mean, as it yielded the best

Table 4
Average results for all color measurement devices using ΔE^*_{ab} color difference equations showing mean and standard deviation (SD) for the *STRESS* index, visual-instrumental agreement, and R-squared values.

Device	STRESS	Visual-instrumental agreement (%)	R^2
eLAB	18	82	0.8
RPC	24	76	0.7
OS	24	76	0.7
PR–670	25	75	0.7
ES-V	25	75	0.7
SSM II	40	60	0.4
Mean	**26**	**74**	**0.7**
sd	7.4	7.4	0.2

Table 5
Results for F-test between different devices using ΔE^*_{ab} color difference equation. Yellow cells in column indicate that a given device performs significantly better than another device in corresponding row. Grey cells indicate significantly worse performance and blue cells indicate no significant difference.

Device	eLAB	RPC	OS	PR-670	ES-V	SSM II
eLAB	1.0	0.563	0.563	0.518	1.000	0.203
RPC	1.778	1.0	1.000	0.922	0.922	0.360
OS	1.778	1.000	1.0	0.922	0.922	0.360
PR-670	1.929	1.085	1.085	1.0	1.000	0.391
ES-V	1.929	1.085	1.000	1.000	1.0	0.391
SSM II	4.938	2.778	2.560	2.560	2.560	1.000

Table 6
Average results for all color measurement devices using the ΔE_{00} difference equation showing standard deviation (SD) for the *STRESS* index, visual-instrumental agreement, and R-squared values.

Device	STRESS	Visual-instrumental agreement (%)	R^2
ES-V	26	74	0.7
PR–670	30	70	0.5
RPC	30	70	0.6
OS	30	70	0.6
eLAB	30	70	0.6
SSM II	32	68	0.5
Mean	**30**	**70**	**0.5**
sd	1.8	1.8	0.05

Table 8
Average results for all color measurement devices using CAM16-UCS, showing standard deviation (SD) for the *STRESS* index, visual-instrumental agreement, and R-squared values.

Device	STRESS	Visual-instrumental agreement (%)	R^2
ES-V	32	68	0.5
SSM II	34	66	0.5
PR–670	36	64	0.4
RPC	39	61	0.4
OS	39	61	0.4
eLAB	39	61	0.4
Mean	**36**	**64**	**0.4**
sd	3.0	3.0	0.1

alignment with the overall measured color differences. From these calculated means, the *STRESS* index for each color difference equation was determined, alongside the evaluation of inter-observer variability [36] which was 45 *STRESS* units on average.

Table 4 illustrates visual-instrumental agreement results for the ΔE^*_{ab} equation, while Table 5 presents F-test outcomes for each device under the ΔE^*_{ab} computation. Visual-instrumental agreement results for the ΔE_{00} equation are displayed in Table 6, with Table 7 showing the corresponding F-test results. Table 8 lists visual-instrumental agreement for CAM16-UCS, and Table 9 presents F-test results for each device under CAM16-UCS computation.

3.2. Visual-instrumental agreement after optimization of color difference equations

Device specific parameters, as optimized by the *fminsearch* function in MATLAB are listed in Table 10 collectively for ΔE', ΔE_{00} and $\Delta \bar{E}$ color difference equations. Table 11 shows the average results (*STRESS* index, visual-instrumental agreement, and R-squared values) for the ΔE' equation while Table 12 further compares the subsequent performance of all devices. In the case of the ΔE_{00} color difference equation, Table 13 and Table 14 present the equivalent improvements and device performance comparison, after the optimization of individual weights (S_L, S_C, S_H). Lastly, Table 15 shows the results for the optimized $\Delta \bar{E}$ color difference formula and Table 16 compares the improved performance among all devices.

3.3. Performance of color difference equations

Further analysis of the performance of the generic and optimized color difference equations for each device using the F-statistic is available in the appendix. The results varied depending on the device, and a comprehensive summary of the performance of each color difference equation across all devices is presented in Fig. 3. The ΔE' equation produced significantly better results more often than other equations under a 95 % significance level.

4. Discussion

While an extensive body of research compares various instruments used to measure tooth color, aiming to establish their relative accuracy [6,7], confusion persists between inter-device agreement and the true definition of colorimetric accuracy [7–10]. Few attempts were made to adequately address the question of accuracy in dental colorimetry by employing a set of calibration standards [37,38]. However, these efforts rely purely on instrumental metrics, overlooking the critical aspect of the congruency between instrumental measurements and visual perception.

The purpose of this study was to investigate the visual-instrumental agreement of six color measurement devices and to optimize three color difference equations based on a visual dataset of color differences obtained from expert observers. The findings shed light on the understudied interplay between device performance and the choice of color difference equations. Statistical analysis, based on the *STRESS* index and associated F-parameter, showed no statistically significant differences in

Table 7
Results for F-test between different devices using ΔE_{00} color difference equation using standard weights (1:1:1) for S_L, S_C, S_H.

Device	ES-V	eLAB	PR-670	RPC	OS	SSM II
ES-V	1.0	0.751	0.751	0.751	0.751	0.522
eLAB	1.331	1.0	1.0	1.0	1.0	0.694
PR-670	1.331	1.0	1.0	1.0	1.0	0.694
RPC	1.331	1.0	1.0	1.0	1.0	0.694
OS	1.331	1.0	1.0	1.0	1.0	0.694
SSM II	1.515	1.138	1.138	1.138	1.138	0.790

Table 9
Results for F-test between different devices using CAM16-UCS color difference equation.

Device	ES-V	SSM II	PR-670	RPC	OS	eLAB
ES-V	1.0	0.886	0.790	0.673	0.673	0.673
SSM II	1.129	1.0	0.892	0.760	0.760	0.760
PR-670	1.266	1.121	1.00	0.852	0.852	0.852
RPC	1.485	1.316	1.174	1.0	1.0	1.0
OS	1.485	1.316	1.174	1.0	1.0	1.0
eLAB	1.485	1.316	1.174	1.0	1.0	1.0

Table 10
Device specific parameters and weights for optimizing color difference equations: S_L, S_C, S_H for the $\Delta E'$, ΔE_{00} and $\Delta\bar{E}$ equations for each of the six devices. Note that the S_L, S_C, S_H parameters are distinct between the different equations.

	$\Delta E'$			ΔE_{oo}			$\Delta\bar{E}$		
Weights	S_L	S_C	S_H	S_L	S_C	S_H	S_L	S_C	S_H
SSM II	0.0	2.6	0.6	3	1	1	1.3	1.0	3.0
PR−670	0.2	2.4	1.0	1.8	0.6	0.9	1.0	1.0	3.0
RPC	0.2	2.4	0.8	2.7	0.8	1.0	1.0	3.0	3.0
OS	0.4	2.1	1.0	2.4	0.8	1.1	1.0	1.8	3.0
eLAB	0.5	1.5	1.1	2.0	0.7	1.1	1.0	1.0	3.0
ES-V	0.5	2.1	0.7	1.6	0.9	1.1	1.0	3.0	3.0

Table 11
Average results for color measurement devices using $\Delta E'$ equation including *STRESS* index, visual-instrumental agreement, and R-squared values.

Device	STRESS	Visual-instrumental agreement (%)	R^2
eLAB	17	83	0.9
RPC	19	81	0.8
OS	20	80	0.8
PR−670	21	79	0.8
ES-V	23	77	0.7
SSM II	24	76	0.7
Mean	**21**	**79**	**0.8**
sd	2.7	2.7	0.1

performance among investigated devices, except for SSM II. Nonetheless, this led to the rejection of the first null hypothesis.

Further analysis yielded unexpected findings regarding the performance of color difference equations. The basic ΔE^*_{ab} formula achieved on average greater agreement with the visual data than the ΔE_{00} equation or CAM16-UCS. One explanation for this might be that this research is restricted to a relatively small gamut (that of tooth color) in color space. However, it is important to acknowledge that, while ΔE^*_{ab} performed well within this limited color gamut, there is ample evidence suggesting that, across a broader spectrum, it may not perform as effectively as the other two metrics. This suggests that the suitability of ΔE^*_{ab} is context-dependent, excelling in specific applications like dental colorimetry but potentially falling short in more expansive color spaces [39].

Many factors may influence color discrimination besides the similarity or dissimilarity of color. The visual task to decide which device and distance measurement correlates best with visually perceived color differences is considerably affected by parametric effects, some of which may be of relevance to clinical dentistry, such as sample edge separation, surface texture, translucency or sample shape and size [40–42]. Failure to account for these may be one reason for low correlation between visual and instrumental color differences [43]. To control for such parametric effects, four hyper realistic phantom models were chosen to resemble the appearance of natural teeth. Considering the large inter-observer variation of 45 *STRESS* units, the observed baseline *STRESS* values for ΔE^*_{ab}, ΔE_{00}, and CAM16-UCS were relatively low, consistent with those reported in other, rigorously controlled visual studies [39,44]. Despite these results, it became evident that further optimization resulted in significantly better visual-instrumental agreement compared to when the generic color difference equations were used (Fig. 3). This led to the rejection of the second null hypothesis.

Johnston [5] described accuracy as an instrument's ability to yield color measurements that align with a reference instrument. However, the criteria for choosing this reference instrument were not specified. More recently [37] this role was assigned to a tele-radiospectrometer of the same type among the investigated devices in the present research. This decision was based on measurements of 240 reference standards provided by a GretagMacbeth DC color checker, yet no direct measure of accuracy, such as the color difference between the reference values and those obtained by the tele-radiospectrometer, was reported. Evaluating inter-device agreement, the performance of Vita Easy Shade and SpectroShade MHT against the nominated gold standard (PR-670) suggested that SpectroShade MHT exhibited the closest congruency with the latter, recommending its use when the gold standard instrument is not

Table 13
Average results for color measurement devices using ΔE_{00} equation, including *STRESS* index, visual-instrumental agreement, and R-squared values.

Device	STRESS	Visual-instrumental agreement (%)	R^2
eLAB	18	82	0.8
PR−670	21	79	0.8
RPC	21	79	0.8
OS	22	78	0.8
ES-V	24	76	0.7
SSM II	27	73	0.6
Mean	**22**	**78**	**0.8**
sd	3.0	3.0	0.1

Table 12
Results for F-test between different devices using optimized $\Delta E'$ equation using custom weights (S_L, S_C, S_H).

Device	eLAB	RPC	OS	PR-670	ES-V	SSM II
eLAB	1.0	0.801	0.903	0.655	0.546	0.502
RPC	1.249	1.0	0.903	0.819	0.682	0.627
OS	1.384	1.108	1.0	0.907	0.756	0.694
PR-670	1.526	1.222	1.103	1.0	0.834	0.766
ES-V	1.830	1.465	1.323	1.200	1.0	0.918
SSM II	1.993	1.596	1.440	1.306	1.089	1.0

Table 14
Results for *F*-test between different devices using ΔE_{00} equation using custom weights S_L, S_C, S_H.

Device	eLAB	PR-670	RPC	OS	ES-V	SSM II
eLAB	1.0	0.735	0.735	0.669	0.563	0.444
PR-670	1.361	1.0	1.000	0.911	0.766	0.605
RPC	1.361	1.000	1.0	0.911	0.766	0.605
OS	1.494	1.098	1.098	1.0	0.840	0.664
ES-V	0.735	1.306	1.306	1.190	1.0	0.790
SSM II	2.250	1.653	1.653	1.506	1.266	1.0

Table 15
Average results for color measurement devices using $\Delta\bar{E}$ equation, including *STRESS* index, visual-instrumental agreement, and R-squared values.

Device	STRESS	Visual-instrumental agreement (%)	R^2
eLAB	19	81	0.8
OS	23	77	0.7
ES-V	24	76	0.7
PR−670	25	75	0.7
RPC	26	74	0.7
SSM II	29	71	0.7
Mean	**24**	**76**	**0.7**
sd	3.3	3.3	0.1

available. In contrast, the present study reveals that the selected color difference equation overshadows variations in device performance. This discrepancy can be attributed to the disparity in methodologies employed, specifically, one based solely on instrumental assessment versus the evaluation of visual-instrumental agreement through the *STRESS*-index, currently regarded as the gold standard in color research [33, 36, 39, 44].

One limitation of this study is the restricted number of sample pairs that were visually scaled, primarily due to the limited availability of distinct color centers of denture teeth in the 3D Master system. Although the 3D Master system offers the broadest range of available shade tabs, which follow a systematic order, its coverage error relative to the gamut of natural tooth color is well documented [45]. Psychometric studies typically fall into one of two categories: either utilizing few expert observers with hundreds of sample pairs to visually scale or employing many observers with few sample pairs [46]. In this study, we opted for the latter, prioritizing hyper-realistic sample pairs resembling the appearance of natural teeth to consider the influence of parametric effects more realistically, even if it meant accepting a smaller population of visually scaled samples. Furthermore, utilizing 26 sample pairs allowed for shorter session durations, preventing eye fatigue or loss of interest. Despite these limitations, the methodology outlined in this study - employing a visual scaling technique for judging the difference between two samples - may offer a pathway for future dental color research by dental researchers, diverging from the conventional approach of measuring identical samples with different devices and comparing multivariate coordinates separately.

Table 16
Results for *F*-test between different devices using $\Delta\bar{E}$ equation using custom weights (S_L, S_C, S_H).

Device	eLAB	OS	ES-V	PR-670	RPC	SSM II
eLAB	1.0	0.682	0.627	0.578	0.534	0.429
OS	1.465	1.0	0.918	0.846	0.783	0.629
ES-V	1.596	1.089	1.0	0.922	0.852	0.685
PR-670	1.731	1.181	1.085	1.0	0.925	0.743
RPC	1.873	1.278	1.174	1.082	1.0	0.804
SSM II	2.330	1.590	1.460	1.346	1.244	1.0

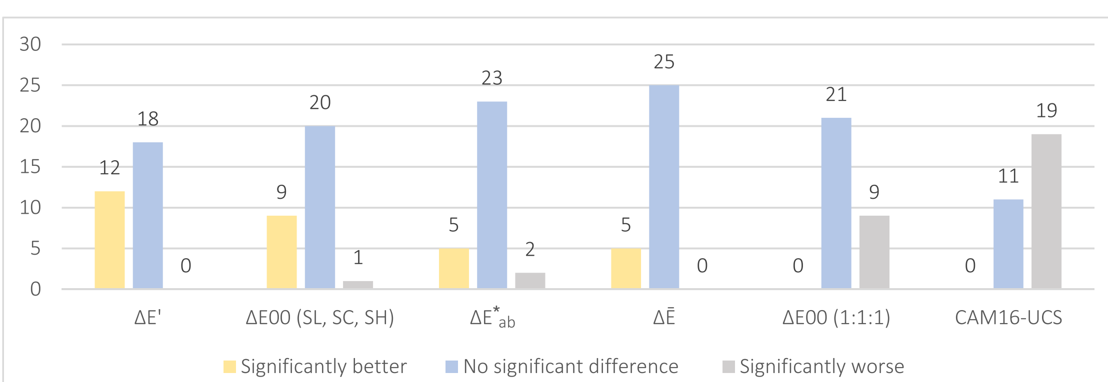

Fig. 3. Average results of performance of color difference equations for each device, counting occurrences where a given equation performed significantly better, worse, or showed no significant difference compared to any other equation, under a 95 % significance level.

5. Conclusion

The findings of our study highlight that discrepancies between visual and instrumental evaluations are primarily influenced by the choice of color difference equation rather than device performance. This suggests that practitioners can significantly enhance color difference prediction by selecting the equation tailored to the specific device in use. Notably, the consistent superiority of the optimized ΔE^*_{ab} equation ($\Delta E'$) across all tested devices underscores its potential for clinical dentistry, providing dental practitioners with a straightforward strategy to improve visual-instrumental agreement in tooth color measurements. This study serves as a foundation for further exploration and refinement of color measurement methodologies, offering valuable insights for both research and practical applications in color science and dentistry.

Acknowledgement

We would like to express our gratitude to all the volunteers who participated in and assisted with this study. In particular we are indebted to: MDT Manuel Gassner, Prof. Olivier Etienne, MDT Johannes Hofmann, Prof. Daniel Edelhoff, Josef Schweiger M.Sc., MDT Christian Vordermayer, Prof. Benedikt Spies, Prof. Jan-Frederik Gueth, Dr. Tobias Graf, Dr. Dominik Kraus, Dr. Norbert Enkling, Prof. Petra Giertmuehlen, Dr. Lea Prott, MDT Patrick Zimmermann, Dr. Yves Mathey, Prof. Martin Schimmel, PD Dr. Alexis Ioannidis, Prof. Lambis Petridis, Assoc Prof. Raelene Sambrook, Dr. Shashwat Bhakta. We would like to thank MDT Michael Zangl for his participation and for providing the hyper realistic phantom models used in this study.

Appendix

This appendix offers a comprehensive analysis of the performance of six color difference equations applied to each of the six color measurement devices tested included in this research. Each table corresponds to a specific device and lists the F-statistic for evaluating the performance of the six color difference equations. These equations include three generic equations and their three optimized versions, adjusted with device specific parameters S_L, S_C, and S_H.

The F-test, conducted to compare the performance among the different equations, is based on a 95 % significance level, with $F_C = 1.955$ and $1/F_C = 0.512$. Yellow cells highlight instances where an equation performs significantly better than another, blue cells denote no significant performance difference, and grey cells signal a significantly poorer performance than the corresponding equation in row.

Table A.1
Performance of color difference equations for eLAB.

Formula	$\Delta E'$	ΔE_{ab}	$\Delta \bar{E}$	ΔE_{00} (S_L, S_C, S_H)	ΔE_{00} (1:1:1)	CAM16-UCS
$\Delta E'$	1.0	0.892	0.801	0.502	0.321	0.190
ΔE_{ab}	1.121	1.0	0.898	0.563	0.360	0.213
$\Delta \bar{E}$	1.249	1.114	1.0	0.627	0.401	0.237
ΔE_{00} (S_L, S_C, S_H)	1.993	1.778	1.596	1.0	0.640	0.379
ΔE_{00} (1:1:1)	3.114	2.778	2.493	1.563	1.0	0.592
CAM16-UCS	5.263	4.694	4.213	2.641	1.690	1.0

Table A.2
Performance of color difference equations for RPC.

Formula	$\Delta E'$	ΔE_{00} (S_L, S_C, S_H)	ΔE_{ab}	$\Delta \bar{E}$	ΔE_{00} (1:1:1)	CAM16-UCS
$\Delta E'$	1.0	0.903	0.627	0.534	0.401	0.237
ΔE_{00} (S_L, S_C, S_H)	1.108	1.0	0.694	0.592	0.444	0.263
ΔE_{ab}	1.596	1.440	1.0	0.852	0.640	0.379
$\Delta \bar{E}$	1.873	1.690	1.174	1.0	0.751	0.444
ΔE_{00} (1:1:1)	2.493	2.250	1.563	1.331	1.0	0.592
CAM16-UCS	4.213	3.803	2.641	2.250	1.690	1.0

Table A.3
Performance of color difference equations for SSM II.

Formula	$\Delta E'$	ΔE_{00} (S_L, S_C, S_H)	$\Delta \bar{E}$	ΔE_{00} (1:1:1)	CAM16-UCS	ΔE_{ab}
$\Delta E'$	1.0	0.790	0.685	0.563	0.498	0.360
ΔE_{00} (S_L, S_C, S_H)	1.266	1.0	0.867	0.712	0.631	0.456
$\Delta \bar{E}$	1.460	1.154	1.0	0.821	0.728	0.526
ΔE_{00} (1:1:1)	1.778	1.405	1.218	1.0	0.886	0.640
CAM16-UCS	2.007	1.586	1.375	1.129	1.0	0.723
ΔE_{ab}	2.778	2.195	1.902	1.563	1.384	1.0

Table A.4
Performance of color difference equations for PR-670.

Formula	$\Delta E'$	ΔE_{00} (S_L, S_C, S_H)	ΔE_{ab}	$\Delta \bar{E}$	ΔE_{00} (1:1:1)	CAM16-UCS
$\Delta E'$	1.0	1.0	0.706	0.706	0.490	0.340
ΔE_{00} (S_L, S_C, S_H)	1.0	1.0	0.706	0.706	0.490	0.340
ΔE_{ab}	1.417	1.417	1.0	1.0	0.694	0.482
$\Delta \bar{E}$	1.417	1.417	1.0	1.0	0.694	0.482
ΔE_{00} (1:1:1)	2.041	2.041	1.440	1.440	1.0	0.694
CAM16-UCS	2.939	2.939	2.074	2.074	1.440	1.0

Table A.5
Performance of color difference equations for OS.

Formula	$\Delta E'$	ΔE_{00} (S_L, S_C, S_H)	$\Delta \bar{E}$	ΔE_{ab}	ΔE_{00} (1:1:1)	CAM16-UCS
$\Delta E'$	1.0	1.0	0.756	0.694	0.444	0.263
ΔE_{00} (S_L, S_C, S_H)	1.0	1.0	0.756	0.694	0.444	0.263
$\Delta \bar{E}$	1.323	1.323	1.0	0.918	0.588	0.348
ΔE_{ab}	1.440	1.440	1.089	1.0	0.640	0.379
ΔE_{00} (1:1:1)	2.250	2.250	1.701	1.563	1.0	0.592
CAM16-UCS	3.803	3.803	2.875	2.641	1.690	1.0

Table A.6
Performance of color difference equations for ES-V.

Formula	$\Delta E'$	ΔE_{00} (S_L, S_C, S_H)	$\Delta \bar{E}$	ΔE_{ab}	ΔE_{00} (1:1:1)	CAM16-UCS
$\Delta E'$	1.0	1.0	0.918	0.846	0.783	0.517
ΔE_{00} (S_L, S_C, S_H)	1.0	1.0	0.918	0.846	0.783	0.517
$\Delta \bar{E}$	1.089	1.089	1.0	0.922	0.852	0.563
ΔE_{ab}	1.181	1.181	1.085	1.0	0.925	0.610
ΔE_{00} (1:1:1)	1.278	1.278	1.174	1.082	1.0	0.660
CAM16-UCS	1.936	1.936	1.778	1.638	1.515	1.0

References

[1] Chu SJ, Trushkowsky RD, Paravina RD. Dental color matching instruments and systems. Review of clinical and research aspects. J Dent 2010;38:e2–16. https://doi.org/10.1016/j.jdent.2010.07.001.

[2] Joiner A, Luo W. Tooth colour and whiteness: a review. J Dent 2017;67:S3–10. https://doi.org/10.1016/j.jdent.2017.09.006.

[3] Macentee M, Lakowski R. Instrumental colour measurement of vital and extracted human teeth. J Oral Rehabil 1981;8:203–8. https://doi.org/10.1111/j.1365-2842.1981.tb00494.x.

[4] Burkinshaw SM. Colour in relation to dentistry. Fundamentals of colour science. discussion 29 Br Dent J 2004;196:33–41. https://doi.org/10.1038/sj.bdj.4810880.

[5] Johnston WM. Color measurement in dentistry. J Dent 2009;37(Suppl 1):e2–6. https://doi.org/10.1016/j.jdent.2009.03.011.

[6] Tabatabaian F, Beyabanaki E, Alirezaei P, Epakchi S. Visual and digital tooth shade selection methods, related effective factors and conditions, and their accuracy and precision: a literature review. J Esthet Restor Dent 2021;33:1084–104. https://doi.org/10.1111/jerd.12816.

[7] Morsy N, Holiel AA. Color difference for shade determination with visual and instrumental methods: a systematic review and meta-analysis. Syst Rev 2023;12:95. https://doi.org/10.1186/s13643-023-02263-9.

[8] Kim-Pusateri S, Brewer J, Davis E, Wee A. Reliability and accuracy of four dental shade-matching devices. J Prosthodont 2009;101:193–9. https://doi.org/10.1016/S0022-3913(09)60028-7.

[9] Chen H, Huang J, Dong X, Qian J, He J, Qu X, et al. A systematic review of visual and instrumental measurements for tooth shade matching. Quintessence Int 2012;43:649–59.

[10] Lehmann KM, Igiel C, Schmidtmann I, Scheller H. Four color-measuring devices compared with a spectrophotometric reference system. J Dent 2010;38:e65–70. https://doi.org/10.1016/j.jdent.2010.07.006.

[11] Igiel C, Weyhrauch M, Wentaschek S, Scheller H, Lehmann KM. Dental color matching: a comparison between visual and instrumental methods. Dent Mater J 2016;35:63–9. https://doi.org/10.4012/dmj.2015-006.

[12] Berns RS. Color and material-appearance measurement. Billmeyer Saltzman'S Princ Color Technol 2019:111–44. https://doi.org/10.1002/9781119367314.ch6.

[13] Hunt RWG, Pointer MR. Precision and accuracy in colorimetry. Meas Colour 2011:197–217. https://doi.org/10.1002/9781119975595.ch9.

[14] Clarke FJJ. High accuracy spectrophotometry at the National Physical Laboratory. J Res Natl Bur Stand 1972;76A:375–403. https://doi.org/10.6028/jres.076A.036.

[15] Robertson AR. Historical development of CIE recommended color difference equations. Color Res Appl 1990;15:167–70. https://doi.org/10.1002/col.5080150308.

[16] McDonald R. Industrial Pass/Fail Colour Matching. Part III-Development of a Pass/Fail Formula for use with Instrumental Measurement of Colour Difference. JSCD 1980;96:486–97. https://doi.org/10.1111/j.1478-4408.1980.tb03544.x.

[17] Luo MR, Cui G, Rigg B. The development of the CIE 2000 colour-difference formula: CIEDE2000. Color Res Appl 2001;26:340–50. https://doi.org/10.1002/col.1049.

[18] B.S. ISO/CIE 11664–6:2014: Colorimetry. CIEDE2000 Colour-difference formula. British Standards Institute; 2014.

[19] Fairchild MD. On the questionable utility of color space for understanding perception. Color Res Appl 2023;48:260–6. https://doi.org/10.1002/col.22853.

[20] Li C, Li Z, Wang Z, Xu Y, Luo MR, Cui G, et al. Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS. Color Res Appl 2017;42:703–18. https://doi.org/10.1002/col.22131.

[21] Melgosa M, Ruiz-Lopez J, Li C, Garcia PA, Della Bona A, Perez MM. Color inconstancy of natural teeth measured under white light-emitting diode illuminants. Dent Mater 2020;36:1680–90. https://doi.org/10.1016/j.dental.2020.10.001.

[22] Hein S, Westland S. Illuminant metamerism between natural teeth and zirconia restorations evaluated with a chromatic adaptation transform. J Prosthodont 2023. https://doi.org/10.1016/j.prosdent.2023.07.035.

[23] Attridge GG, Pointer MR. Some aspects of the visual scaling of large colour differences-II. Color Res Appl 2000;25:116–22. https://doi.org/10.1002/(SICI)1520-6378(200004)25:2<116::AID-COL6>3.0.CO;2-9.

[24] Luo MR, Hunt RWG. Testing colour appearance models using corresponding-colour and magnitude-estimation data sets. Color Res Appl 1998;23:147–53. https://doi.org/10.1002/(SICI)1520-6378(199806)23:3<147::AID-COL6>3.0.CO;2-Q.
[25] Pan Q, Westland S. Comparative evaluation of color differences between color palettes. Soc Imaging Sci Technol 2018.
[26] Paravina RD, Pérez MM, Ghinea R. Acceptability and perceptibility thresholds in dentistry: a comprehensive review of clinical and research applications. J Esthet Restor Dent 2019;31:103–12. https://doi.org/10.1111/jerd.12465.
[27] Shen S, Berns RS. Evaluating color difference equation performance incorporating visual uncertainty. Color Res Appl 2009;34:375–90. https://doi.org/10.1002/col.20521.
[28] Berns RS, Alman DH, Reniff L, Snyder GD, Balonon-Rosen MR. Visual determination of suprathreshold color-difference tolerances using probit analysis. Color Res Appl 1991;16:297–316. https://doi.org/10.1002/col.5080160505.
[29] CIE 015:2018. Colorimetry, 4th ed. Technical Report. CIE Central Bureau, Vienna. 2018.
[30] Hein S, Modrić D, Westland S, Tomeček M. Objective shade matching, communication, and reproduction by combining dental photography and numeric shade quantification. J Esthet Restor Dent 2021;33:107–17. https://doi.org/10.1111/jerd.12641.
[31] P.D. ISO/TR 28642:2016: Dentistry. Guidance on colour measurement. British Standards Institute; 2016.
[32] Westland S, Ripamonti C, Cheung V. Computational colour science using MATLAB. Second edition. ed. Chichester. Wiley,; 2012.
[33] García PA, Huertas R, Melgosa M, Cui G. Measurement of the relationship between perceived and computed color differences. JOSA A 2007;24:1823–9. https://doi.org/10.1364/JOSAA.24.001823.
[34] Huang M, Cui G, Melgosa M, Sanchez-Maranon M, Li C, Luo MR, et al. Power functions improving the performance of color-difference formulas. Opt Express 2015;23:597–610. https://doi.org/10.1364/OE.23.000597.
[35] CIE 217:2016. Recommended Method for Evaluating the Performance of Colour-Difference Formulae. CIECentral Bureau, Vienna. 2016.
[36] Melgosa M, García PA, Gómez-Robledo L, Shamey R, Hinks D, Cui G, et al. Notes on the application of the standardized residual sum of squares index for the assessment of intra- and inter-observer variability in color-difference experiments. JOSA A 2011;28:949–53. https://doi.org/10.1364/JOSAA.28.000949.
[37] Akl MA, Sim CPC, Nunn ME, Zeng LL, Hamza TA, Wee AG. Validation of two clinical color measuring instruments for use in dental research. J Dent 2022;125:104223. https://doi.org/10.1016/j.jdent.2022.104223.
[38] Gozalo-Diaz DJ, Lindsey DT, Johnston WM, Wee AG. Measurement of color for craniofacial structures using a 45/0-degree optical configuration. J Prosthet Dent 2007;97:45–53. ⟨http://doi.org/10.1016/j.prosdent.2006.10.013⟩.
[39] Luo MR, Xu Q, Pointer M, Melgosa M, Cui G, Li C, et al. A comprehensive test of colour-difference formulae and uniform colour spaces using available visual datasets. Color Res Appl 2023;48:267–82. https://doi.org/10.1002/col.22844.
[40] Xu Q, Shi K, Luo MR. Parametric effects in color-difference evaluation. Opt Express 2022;30:33302–19. https://doi.org/10.1364/OE.462628.
[41] Parametric effects in colour-difference evaluation, CIE Publ. No. 101, Central Bureau of the CIE, Vienna, 1993. Col Res Appl. 1993;18:289–. http://doi.org/10.1002/col.5080180415.
[42] Guan S-S, Luo MR. Investigation of parametric effects using small colour differences. Color Res Appl 1999;24:331–43. https://doi.org/10.1002/(SICI)1520-6378(199910)24:5<331::AID-COL5>3.0.CO;2-9.
[43] Berns RS. Fourth edition ed. Newark: John. Billmeyer and Saltzman's Principles of Color Technology. Wiley & Sons, Incorporated,; 2019. https://doi.org/10.1002/9781119367314.
[44] Melgosa M, Huertas R, Berns RS. Performance of recent advanced color-difference formulas using the standardized residual sum of squares index. JOSA A 2008;25:1828–34. https://doi.org/10.1364/JOSAA.25.001828.
[45] Ruiz-López J, Perez MM, Lucena C, Pulgar R, López-Toruño A, Tejada-Casado M, et al. Visual and instrumental coverage error of two dental shade guides: an in vivo study. Clin Oral Invest 2022;26:5961–8. https://doi.org/10.1007/s00784-022-04556-0.
[46] Pointer MR, Attridge GG. Some aspects of the visual scaling of large colour differences. Color Res Appl 1997;22:298–307. https://doi.org/10.1002/(SICI)1520-6378(199710)22:5<298::AID-COL3>3.0.CO;2-S.