ARTICLE IN PRESS

Dental Materials xxx (xxxx) xxx

Contents lists available at ScienceDirect

Dental Materials

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

Evaluating visual thresholds and color metrics in dental research: An exploratory study

Sascha Hein [a,*], Michael Zangl [b,c], Tobias Graf [d], Kirstin Vach [e], Jan-Frederik Güth [d], Stephen Westland [a]

[a] *School of Design, University of Leeds, Leeds, UK*
[b] *Private Dental Laboratory, Zahntechnik-Cham GmbH, Waldschmidtstraße 11, Cham 93413, Germany*
[c] *Department of Prosthodontics, Faculty of Dentistry, Domstr. 8, 17489 Greifswald, Germany*
[d] *Department of Prosthetic Dentistry, Center for Dentistry and Oral Medicine (Carolinum), Goethe University Frankfurt am Main, Theodor-Stern-Kai 7, 60596 Frankfurt am Main, Germany*
[e] *Hannover Medical School (MHH), Department of Conservative Dentistry, Periodontology and Preventive Dentistry, Carl-Neuberg-Str.1, 30625 Hannover, Germany*

ARTICLE INFO

Keywords:
Tooth color measurement
Color difference equations
CIELAB
CIEDE2000
Visual thresholds
STRESS index
Visual scaling
Magnitude estimation
VIAS

ABSTRACT

Objectives: This study aimed to evaluate perceptibility (PT) and acceptability thresholds (AT) for multiple color measurement devices and assess the performance of three color difference equations (ΔE^*_{ab}, ΔE_{00}, and ΔE_{94}) using a visual dataset from expert observers.
Methods: A visual dataset previously published was extended by adding the x-rite MetaVue spectrophotometer and ΔE_{94} to the analysis. Visual scaling was performed on 26 sample pairs of teeth using magnitude estimation. Observers answered PT and AT questions to determine thresholds. Threshold estimation was conducted using a model-free method, and device performance was analyzed using the standardized residual sum of squares (*STRESS)* index and visual instrument agreement scale (*VIAS*).
Results: The PT and AT thresholds varied across devices and color difference equations. For ΔE_{00}, *STRESS* values ranged from 23 to 32 (mean 29, sd 2.9), with *VIAS* scores between 68 % and 77 % (mean 71 %, sd 2.9). ΔE_{94} showed higher *STRESS* values (24–42, mean 34, sd 5.5) and lower *VIAS* scores (58–76 %, mean 66 %, sd 5.5). ΔE^*_{ab} demonstrated excellent visual-instrumental agreement with *STRESS* values from 18 to 36 (mean 24, sd 5.9) and lower *VIAS* scores (82–64 %, mean 76 %, sd 5.9) outperforming ΔE_{94} and ΔE_{00}. The x-rite MetaVue achieved excellent results under controlled conditions but it is unsuitable for clinical research due to its design.
Significance: This study highlights the variability in PT and AT across devices, suggesting the need for device-specific thresholds. It also demonstrates the effectiveness of ΔE^*_{ab} in dental colorimetry compared to more complex color difference metrics

1. Introduction

Color technology in industry and business has traditionally focused on quality assessment, with particular emphasis on determining whether a pair of samples match [1]. Similarly, in dental research, tooth color assessment is a frequent subject of investigation [2–6], reflecting its critical role in patient satisfaction [7]. Shade matching a single anterior tooth with a restoration is often crucial, but differences in esthetic expectations, rising demands, and the challenges of accurate color determination frequently result in esthetic failures [8].

Instrumental measurements should align with visual perception, ensuring that calculated color differences reflect those observed by individuals [9]. Ideally, a restoration should perfectly and unconditionally match its natural counterpart, but the complexity of tooth color appearance [10] can make this an unattainable ideal. To address this, there are two general types of visual assessments—perceptibility and acceptability—applicable not only in clinical dentistry but across all industries involved in color management [11,12]. Industry-specific needs are addressed through dedicated psychophysical experiments designed to estimate appropriate thresholds. A sigmoidal transformation is applied to predict the computed color difference at which 50 % of the expert observer population can perceive a color difference between a

* Correspondence to: Graduate School of Color Science and Technology, School of Design, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK.
E-mail addresses: sdsch@leeds.ac.uk (S. Hein), t.graf@med.uni-frankfurt.de (T. Graf), gueth@med.uni-frankfurt.de (J.-F. Güth).

https://doi.org/10.1016/j.dental.2025.04.006
Received 25 November 2024; Received in revised form 17 February 2025; Accepted 15 April 2025

test sample pair [13]. For this so-called *50/50 perceptibility threshold* (PT), it is not uncommon for sample pairs with color differences only slightly above or equal to the PT to be judged as unacceptable, particularly when whitish-pale samples are visually evaluated by expert observers [14]. In many industries this can create significant practical challenges, as such PT values are typically small and often exceed production tolerances. To address this, a so-called *commercial factor* is often introduced to establish a more practical *50/50 acceptability threshold* (AT) [15].

PT and AT thresholds for use in dentistry were established by Paravina et al. [16] using flat, uniformly colored (monochromatic) ceramic samples visually scaled by a mixed population, including dental practitioners, technicians, auxiliaries, students, and laypeople. Samples were measured with a spectroradiometer under consistent viewing conditions, and color differences were calculated using the CIE 1976 Euclidean formula (ΔE^*_{ab}) and the more modern CIEDE2000 (ΔE_{00}) metric, following CIE recommendations for small color differences [17]. The average PT and AT values for both metrics, PT (1.2 ΔE^*_{ab} / 0.8 ΔE_{00}) and AT (2.7 ΔE^*_{ab} / 1.8 ΔE_{00}), were subsequently adopted by the International Organization for Standardization (ISO/TR 28642:2016, Dentistry – Guidance on Colour Measurement) [18] and have since become a benchmark for numerous scientific investigations, where they tended to be applied regardless of the instrument or illumination geometry used [2,19]. However, these thresholds have often been applied without considering the potential impact of differences in instrumentation and measurement geometry [20,21].

Instrumental shade measurement in dentistry has seen renewed interest with the introduction of new devices and increasing scientific inquiry into their clinical and research applications [22–24]. Recent efforts have focused on estimating visual-instrumental agreement using the standardized residual sum of squares (*STRESS*) index and evaluated through the recently termed *Visual Instrument Agreement Scale* (*VIAS*) [25]. A multi-center study demonstrated that device performance in *VIAS* assessments was strongly influenced by the choice of color difference equation [26]. One device, the 'optishade' (Smile Line, Switzerland), uniquely employs the ΔE_{94} metric, an unusual choice given that it has been largely superseded by ΔE_{00} in general color science applications. However, its increasing popularity among dental practitioners, due to its ease of use in clinical settings, raises the question of whether ΔE_{94} remains a viable alternative for tooth color measurement. Moreover, reliable PT and AT thresholds for ΔE_{94} have not yet been established, and its visual-instrumental agreement remains unexplored. Investigating its suitability is therefore important, particularly if it offers comparable or superior performance to ΔE_{00} in the specific context of tooth color measurement.

Another device, the x-rite 'MetaVue' imaging spectrophotometer, has been highlighted in the literature [27] for its suitability for measuring diffusely scattering media without causing edge loss [28]. However, its effectiveness in dental research remains unverified, warranting further investigation. Establishing robust PT and AT thresholds is critical, as CIELAB values are well known to be device-dependent [3,29]. The human eye remains the final arbiter in determining whether a restoration is acceptable based on its color match. Nevertheless, visual thresholds serve as essential benchmarks for standardizing instrumental color measurements, ensuring reproducibility and facilitating effective communication between clinicians and dental laboratories.

Therefore, the aim of this exploratory study was to shed light on whether the PT and AT values recommended for color measurement in dentistry are universally applicable across different devices and illumination geometries, whether expert thresholds differ from general thresholds, potentially indicating the effects of a commercial factor, and to what extent ΔE_{94} PT and AT thresholds align with or differ from ΔE_{00}, given that the latter metric evolved from the former. Additionally, the study aimed to evaluate whether ΔE_{94} is justified by superior *STRESS* and *VIAS* performance compared to other metrics and to investigate the suitability of the MetaVue spectrophotometer for dental color measurement.

2. Material and methods

The present study utilizes a visual dataset previously acquired; details about how tooth colors were measured and how the visual experiment was conducted have been previously published [26]. Only an abridged version of the methodology is presented here, with additional elements unique to this study highlighted, including the inclusion of one extra device and one additional color difference equation.

2.1. Selection of expert observers and sites

Ethical approval (LTDESN-196) was obtained, allowing recruitment of 154 expert observers, comprising dental practitioners and dental technicians with experience in restorative dentistry. Observers passed the Ishihara test for color vision deficiency and were recruited across 16 professional settings, including dental schools, dental laboratories, and private practices.

2.2. Visual scaling technique

To examine visual color differences (ΔV) between pairs of teeth, the magnitude estimation (ME) technique was applied, where observers assessed color differences between maxillary central incisors. They were asked to rate the color match on a scale from 0 % to 100 %, with 0 % indicating the worst possible match and 100 % indicating a perfect match. Observers were also asked two specific questions to determine PT and AT:

1. *"Can you see a color difference between the two maxillary centrals?"*
2. *"Would you accept this color difference if this were your patient?"*

2.3. Visually scaled samples and experimental setup

Previous studies on visual scaling of tooth-colored samples often used simplified samples, such as monochromatic ceramic discs [30] or other configurations, designed to minimize the influence of parametric effects [31]. In this study, a different approach was taken by employing custom-made, hyper-realistic phantom models that closely resemble the appearance of natural teeth, aiming to create a lifelike visual context for observers. Four color centers within the CIELAB color space were identified, representing one base shade, with 5–7 exchangeable teeth per model, resulting in a total of 26 visually scaled sample pairs with controlled variations in hue and chroma, all with color differences under 5 ΔE^*_{ab} units. Observations were conducted at a distance of 35–50 cm against a neutral grey background under simulated daylight at 6500 K.

2.4. Instrumental measurement of sample pairs

Color measurements were obtained using devices frequently cited in dental literature, listed in Table 1. An additional device, the x-rite MetaVue, was included due to recent positive references in the literature [27]. Sample pairs were mostly measured with devices designed for dentistry, with each device using a consistent measurement protocol and illumination geometry.

2.5. Computation of inter-instrument variability

To evaluate inter-instrument variability in color measurements, pairwise subtractions of color differences were performed between the tested instruments. Color differences were computed under Illuminant D65 for the CIE 1931 standard colorimetric observer, using the CIE D65 reference white (X = 95.047, Y = 100.000, Z = 108.883). Three color difference equations were employed: ΔE^*_{ab}, ΔE_{00}, and, additionally, ΔE_{94}, as it is used by the Optishade dental colorimeter which has

Table 1
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	Spectroradiometer
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	Spectrophotometer
MetaVue (MetaVue)	X-Rite Inc. USA	Grand Rapids, MI, USA	Imaging spectrophotometer

recently gained attention in dental research [24,32,33]. For instruments *i* and *j*, the pairwise difference was calculated as:

$$\Delta E_{difference,ij} = \Delta E_i - \Delta E_j$$

where $\Delta E_{\text{difference, ij}}$ represents the variability in color differences between instruments *i* and *j*, and ΔE_i and ΔE_j denotes the color differences computed for instruments *i* and *j*, respectively. This calculation was repeated for all pairwise comparisons across the instruments for each color difference equation ΔE^*_{ab}, ΔE_{00}, and ΔE_{94}.

2.6. Computation of visual thresholds

Visual thresholds were determined based on observer responses to the PT and AT questions. The computation of thresholds used a nonparametric, Model-Free Estimation Technique developed by Zychaluk & Foster [34], which uses local linear fitting. This approach does not assume a specific parametric model for the psychometric function but rather relies on a smoothness assumption, providing threshold estimates that adapt to the response distribution [35]. This method is particularly effective as it mitigates potential biases from model misspecifications, yielding robust and consistent threshold values.

2.7. Statistical analysis

Data evaluation was conducted using MATLAB (R2024b) with a specialized color science toolbox, and statistical testing was performed using STATA (Version 17.0, College Station, TX, USA) with a significance level of 5 %. Linear mixed models, incorporating sample pair as a random effect, were applied to test for inter-instrument variability across devices for each of the three color difference (ΔE_{00}, ΔE_{94}, and ΔE^*_{ab}). The following pairwise comparisons were not corrected for multiple testing due to the exploratory nature of the study. Both, the *STRESS* index and *VIAS* were used to evaluate device performance. Small *STRESS* values indicate high visual-instrumental agreement, while *VIAS* is calculated as 100 - *STRESS*, meaning that higher *VIAS* values correspond to greater visual-instrumental agreement.

3. Results

3.1. Inter-instrument variability

The results for inter-instrument variability, calculated using the ΔE_{00}, ΔE_{94}, and ΔE^*_{ab} color difference equations, are shown in Figs. 1, 2, and 3, respectively. Among the 21 different device pair comparisons, 12 combinations exhibited significant color differences ($p < 0.05$) for both the ΔE_{00} and ΔE_{94} color difference equations. For the ΔE^*_{ab} equation, six device pairs demonstrated statistically significant differences. Individual color differences between all devices for each of three color

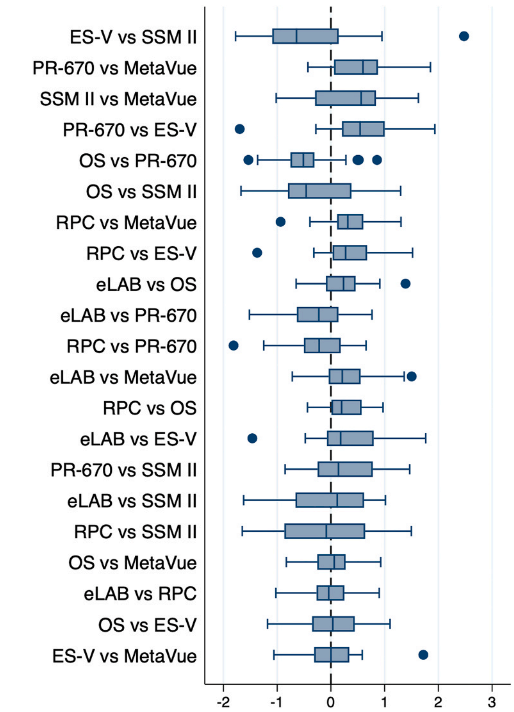

Fig. 1. Boxplot showing pairwise comparisons of inter-instrument variability calculated using the ΔE_{00} color difference equation. Each box represents the ΔE_{00} differences between the first and second named device, sorted by increasing absolute median values from bottom to top. Dots represent outliers. The dashed line indicates no difference between devices. Among the 21 device pair comparisons, 12 pairs showed statistically significant differences ($p < 0.05$).

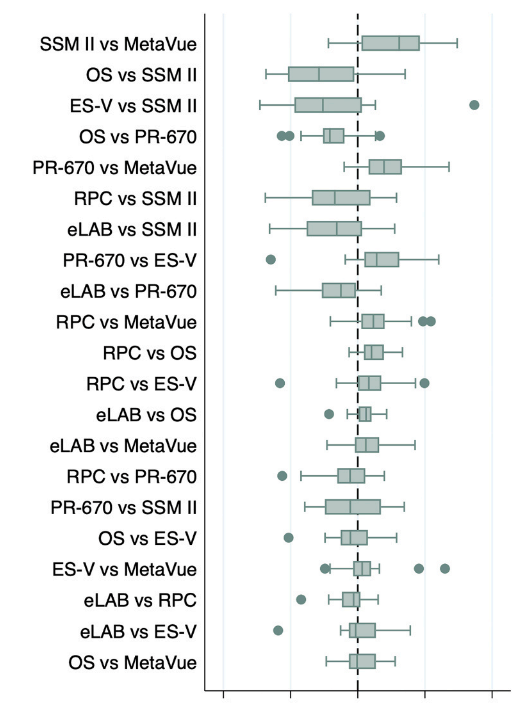

Fig. 2. Boxplot showing pairwise comparisons of inter-instrument variability calculated using the ΔE_{94} color difference equation. Each box represents the ΔE_{94} differences between the first and second named device, sorted by increasing absolute median values from bottom to top. Dots represent outliers. The dashed line indicates no difference between devices. Out of 21 device pair comparisons, 12 pairs exhibited statistically significant differences ($p < 0.05$).

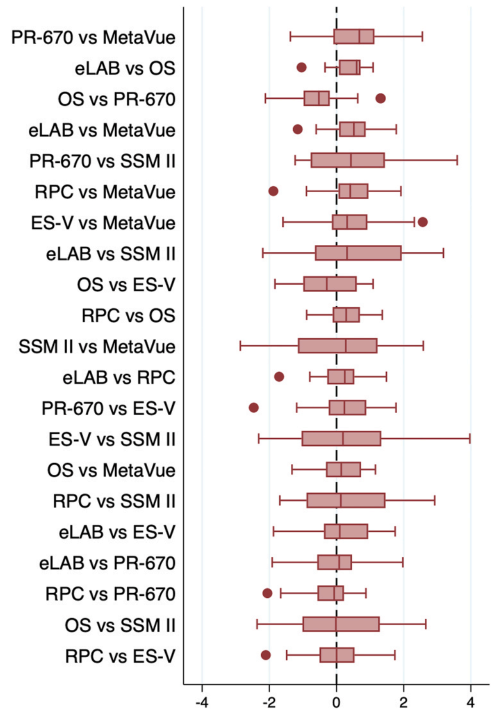

Fig. 3. Boxplot showing pairwise comparisons of inter-instrument variability calculated using the ΔE^*_{ab} color difference equation. Each box represents the ΔE^*_{ab} differences between the first and second named device, sorted by increasing absolute median values from bottom to top. Dots represent outliers. The dashed line indicates no difference between devices. Six out of 21 device pairs showed statistically significant differences ($p < 0.05$).

difference equations are included in appendix A.

Since visual thresholds calculated by model-free estimation, are sensitive to variations of ΔE^* values per device and color difference equation, equivalence class partitioning was applied. This approach grouped devices into clusters with no significant differences in color differences among members as shown in Fig. 4.

3.2. Visual thresholds

The results for the 50/50 PT and AT for each device, calculated using the ΔE_{00}, ΔE_{94}, and ΔE^*_{ab} color difference equations, are summarized in Tables 2, 3, and 4, respectively. These tables also include associated statistical metrics, including Standard Error (SE), the 95 % confidence interval (CI) low and high values, as well as the average values and standard deviations across all devices. Results are reported by equivalence class groups, with the mean for each group, the grand mean, and standard deviation (SD) provided. The average PT values were 0.8 for ΔE_{00}, 0.9 for ΔE_{94}, and 1.2 for ΔE^*_{ab}, while the AT values were 1.8 for both ΔE_{00} and ΔE_{94}, and 2.8 for ΔE^*_{ab}. Variations in PT and AT values were observed across different devices and color difference equations.

3.3. STRESS index and VIAS score

The results for the *STRESS* index, *VIAS* and *R*-squared for each color difference equation (ΔE_{00}, ΔE_{94}, and ΔE^*_{ab}) are presented in Tables 5, 6, and 7, respectively. Each table also includes the average values across all devices and their standard deviations. Individual device performances are detailed in Tables B1, B2, and B3 in the appendix.

4. Discussion

This exploratory study aimed to evaluate the applicability of PT and AT thresholds across devices with differing designs and illumination geometries, the potential influence of a commercial factor on expert thresholds, and the relationship between ΔE_{94} and ΔE_{00} thresholds. Additionally, it assessed whether ΔE_{94} offers superior *STRESS* and *VIAS* performance and the suitability of the MetaVue spectrophotometer in dental research.

The established PT and AT thresholds were derived from psychophysical experiments that employed the Takagi-Sugeno-Kang (TSK) Fuzzy Approximation technique [16]. The experimental setup, designed to provide a controlled and repeatable testing environment, utilized flat, square ceramic samples to represent teeth and an opaque pink barrier to approximate the presence of gingiva. While this approach offers a structured framework for assessing color difference perception, it represents a highly simplified model of the anatomical and optical complexities found in the oral cavity. In contrast, the present study employed highly realistic phantom models to better replicate clinical conditions and included expert observers exclusively, whereas previous studies incorporated a mixed population.

The present study employed the Model-Free Estimation Technique by Zychaluk & Foster [34,35], a non-parametric method that approximates psychometric functions using local linear fitting with kernel smoothing, adapting dynamically to data through cross-validation. This approach was chosen for its statistical rigor and open accessibility, as it

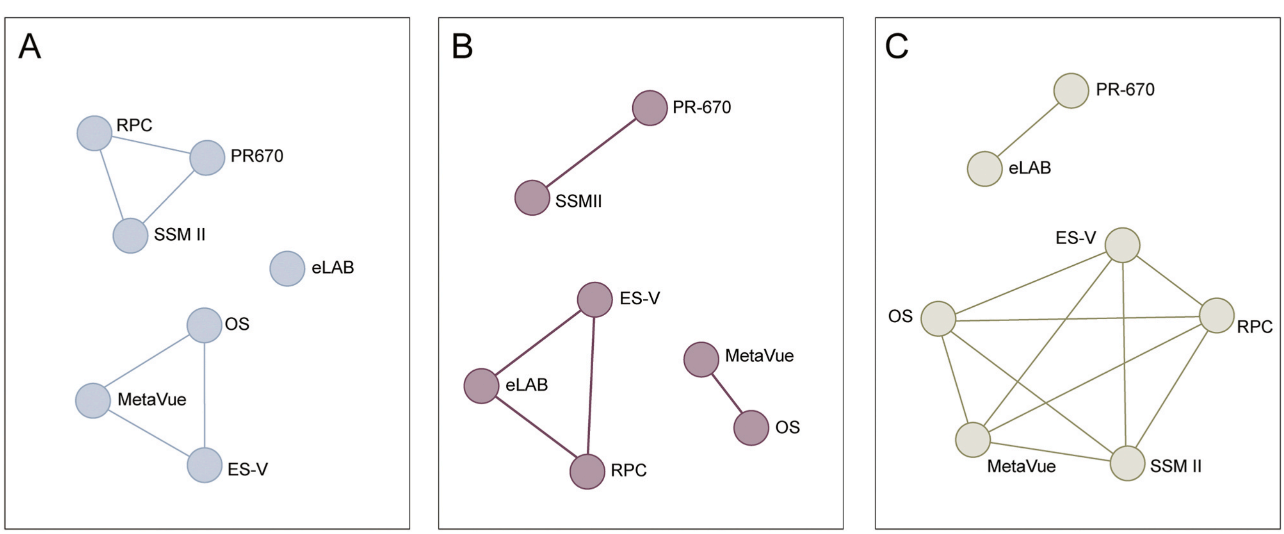

Fig. 4. Equivalence class plots for ΔE_{00} (A), ΔE_{94} (B), and ΔE^*_{ab} (C). Each connected component represents devices with no significant differences among them.

Table 2
Summary of 50/50 PT and AT values for each device using the ΔE_{00} color difference equation. Statistical metrics include Standard Error (SE), 95 % confidence interval (CI) low and high values, average values, and standard deviations across all devices. Results are reported by equivalence class groups, with the mean for each group, the grand mean, and standard deviation (SD) also provided.

Device	PT(ΔE_{00})	SE	95 % CI low	95 % CI high	AT(ΔE_{00})	SE	95 % CI low	95 % CI high
ES-V	0.9	0.021	0.849	0.933	1.6	0.021	1.628	1.706
MetaVue	0.8	0.028	0.713	0.82	1.7	0.021	1.643	1.731
OS	0.6	0.028	0.596	0.729	1.7	0.030	1.681	1.800
Group Mean	0.8				1.7			
RPC	0.7	0.05	0.55	0.742	1.8	0.033	1.855	1.972
PR−670	0.8	0.055	0.69	0.904	2.1	0.036	2.040	2.192
SSM II	0.9	0.001	0.911	0.911	1.8	0.054	1.627	1.867
Group Mean	0.9				1.9			
eLAB	0.7	0.039	0.612	0.778	1.9	0.025	1.814	1.920
Grand Mean	0.8				1.8			
sd	0.111				0.166			

Table 3
Summary of 50/50 PT and AT values for each device using the ΔE_{94} color difference equation. See Table 2 for detailed metrics and grouping information.

Device	PT(ΔE_{94})	SE	95 % CI low	95 % CI high	AT(ΔE_{94})	SE	95 % CI low	95 % CI high
RPC	0.9	0.039	0.814	0.955	1.8	0.034	1.865	1.989
eLAB	0.6	0.048	0.511	0.703	1.8	0.026	1.775	1.882
ES-V	1.0	0.018	1.004	1.072	1.7	0.022	1.725	1.806
Group Mean	0.8				1.8			
PR−670	0.8	0.031	0.751	0.864	2.1	0.052	2.153	2.319
SSM II	1.3	0.001	1.255	1.255	1.7	0.126	1.407	1.970
Group Mean	1.1				1.9			
OS	0.7	0.023	0.612	0.694	1.5	0.026	1.545	1.648
MetaVue	0.7	0.027	0.711	0.814	1.7	0.025	1.669	1.757
Group Mean	0.7				1.6			
Grand Mean	0.9				1.8			
sd	0.230				0.183			

Table 4
Summary of 50/50 PT and AT values for each device using the ΔE^*_{ab} color difference equation. See Table 2 for detailed metrics and grouping information.

Device	PT(ΔE^*_{ab})	SE	95 % CI low	95 % CI high	AT(ΔE^*_{ab})	SE	95 % CI low	95 % CI high
OS	0.9	0.044	0.875	1.046	2.8	0.036	2.711	2.859
SSM II	1.3	0.000	1.345	1.345	2.5	0.072	2.320	2.594
ES-V	1.4	0.042	1.337	1.504	2.7	0.030	2.715	2.824
MetaVue	1.2	0.001	1.158	1.158	2.6	0.038	2.545	2.680
RPC	1.0	0.075	0.855	1.123	2.8	0.047	2.822	3.009
Group Mean	1.2				2.7			
eLAB	1.1	0.054	1.032	1.242	3.1	0.043	3.007	3.152
PR−670	1.2	0.074	1.041	1.336	3.2	0.047	3.105	3.296
Group Mean	1.2				3.1			
Grand Mean	1.2				2.8			
sd	0.173				0.251			

Table 5
STRESS index, *VIAS* scores, and R^2 values for each device using the ΔE_{00} color difference equation, with devices listed in order of *STRESS* values from low to high. The table includes mean and standard deviation (SD) values for all devices, providing an overview of performance across devices in terms of agreement and variance.

Device	*STRESS*	*VIAS*	R^2
MetaVue	23	77	0.7
ES-V	26	74	0.7
OS	29	71	0.6
eLAB	30	70	0.6
PR−670	30	70	0.5
RPC	30	70	0.6
SSM II	32	68	0.5
Mean	29	71	0.6
SD	2.9	2.9	0.1

Table 6
STRESS index, *VIAS* scores, and R^2 values for each device using the ΔE_{94} color difference equation. See Table 5 for detailed metrics and ordering information.

Device	*STRESS*	*VIAS*	R^2
MetaVue	24	76	0.7
eLAB	33	67	0.5
ES-V	33	67	0.5
OS	34	66	0.5
PR−670	34	66	0.4
RPC	37	63	0.4
SSM II	42	58	0.2
Mean	34	66	0.5
SD	5.5	5.5	0.2

can be implemented using freely available software such as MATLAB. In contrast, TSK Fuzzy Approximation relies on fuzzy logic with Gaussian membership functions and rule-based inference. While the TSK approach offers interpretability through linguistic rules, it requires careful selection of membership functions, whereas the Model-Free

Table 7
STRESS index, *VIAS* scores, and R^2 values for each device using the ΔE^*_{ab} color difference equation. See Table 5 for detailed metrics and ordering information.

Device	*STRESS*	*VIAS*	R^2
eLAB	18	82	0.8
MetaVue	20	80	0.8
OS	22	78	0.8
RPC	24	76	0.7
PR−670	25	75	0.7
ES-V	25	75	0.7
SSM II	36	64	0.3
Mean	24	76	0.7
SD	5.9	5.9	0.2

Estimation Technique is purely statistical, making it particularly well-suited for psychometric function estimation. While computational differences may introduce small numerical variations, the overall results are expected to be comparable. This is reflected in the present study, where the average PT and AT values align closely with those reported by Paravina et al., further supporting the robustness of both estimation techniques and indicating their comparable performance.

In highly controlled psychophysical experiments, expert observers, such as colorists or quality control assessors, are often preferred because their trained visual acuity and experience result in lower inter- and intra-observer variability, producing more robust data and minimizing measurement noise [36–39]. However, individual quality assessments, even by experts, are subject to error, as demonstrated in other industries where visual pass/fail decisions play a critical role. Studies on professional shade passers in large-scale production environments have shown that, on average, 17 % of visual judgments were incorrect, evenly split between false acceptances and false rejections. The consequences of such misjudgments include unnecessary remakes, increased production costs, and reputational damage [39]. While aggregated expert assessment can significantly reduce visual misjudgments, it is impractical for routine application. The most effective solution is instrumental color measurement, guided by thresholds derived from aggregated expert assessments [40].

Similarly, dental practitioners must independently assess the color match of restorations without the benefit of aggregated expert judgments, introducing uncertainty and increasing the risk of unnecessary remakes or acceptance of suboptimal restorations. Visual thresholds based on expert assessments help mitigate this risk by providing an objective benchmark for decision-making. A considerable body of research on visual threshold estimation in dentistry has been reviewed by Paravina et al. [41], revealing substantial variation in reported thresholds across different devices, and research methodologies ($PT_{mean} = 1.6\ \Delta E^*_{ab}$ (SD 0.7); $AT_{mean} = 3.2\ \Delta E^*_{ab}$ (SD 1.0)). The majority of these studies relied on non-expert observers to ensure practical execution and to simulate a general patient population (i.e., laypersons). The resulting higher PT and AT values suggest a commercial factor, as non-expert observers tend to be less critical in their assessments [30,42,43]. This effect was also observed in the original work by Paravina et al. [16], where dentists and dental technicians demonstrated notably lower thresholds compared to dental students, auxiliaries, and laypersons.

In addition, a robust visual dataset obtained from expert observers is essential for detecting device-dependent threshold differences, which are likely to be small and easily obscured by noise in the data. Statistical testing revealed frequent and significant inter-device variability in PT and AT values, depending on the device and the color difference metric used, differences that may have otherwise been missed with a more variable observer population.

To better understand these variations, equivalence class partitioning was applied to group devices with no significant inter-device variability. Within these groups, PT and AT values remained relatively consistent, supporting the idea that thresholds can be applied uniformly within a group. However, differences between groups highlight the need for device-specific or group-specific thresholds, suggesting that a single set of values cannot be uniformly applied across all devices without accounting for their design, measurement geometry, and equivalence class groupings

The use of ΔE_{94}, a metric that has been superseded by ΔE_{00} [17], may be seen as an unusual choice, which sparked interest in investigating its effectiveness. While the findings do not align with the conclusions of Rizzi et al. [44], which suggested that ΔE_{94} performs better than ΔE_{00} in aligning with visual perception, the broader notion that simpler color difference metrics may perform better for tooth colors is supported. In this study, ΔE_{94} did not outperform ΔE_{00}, which is consistent with the historical development of ΔE_{00} as an evolution of ΔE_{94} to address discrepancies in regions of CIELAB beyond the gamut of natural tooth colors [45]. Both metrics produced nearly identical results for visual thresholds but ΔE_{00} outperformed ΔE_{94} notebaly as evaluated by *STESS* and *VIAS* metrics.

The findings reaffirm the effectiveness and computational simplicity of ΔE^*_{ab}, which delivered superior results without the added complexity of more modern equations. This can be attributed to the location of natural tooth colors within a region of CIELAB [46] where color difference ellipsoids are small and spherical, as demonstrated by Luo and Riggs [47], and where the assumption of perceptual uniformity reasonably holds.

Another relatively new device mentioned in scientific research, the x-rite MetaVue, demonstrated excellent performance, as indicated by *STRESS* and *VIAS* results. However, while this instrument may be beneficial for in-vitro investigations, it is not suitable for clinical research due to its design.

The present study acknowledges the importance of device-specific thresholds and recommends replacing the common practice of applying a single set of values uniformly across all color measurement devices, even though these differences may appear small. While the study aimed to simulate real-life clinical conditions using hyper-realistic phantom models, the results require clinical validation. Although the use of expert observers was intended to minimize intra- and inter-observer variability, future research should investigate the impact of lay observers to assess potential commercial factor effects. Further work is also needed to compare the performance of TSK fuzzy approximation against the Model-Free Estimation Technique, which would require open-source access to specific TSK fuzzy parameters (e.g., membership functions) to better elucidate visual color difference perception in dentistry.

5. Conclusions

Within its limitations, the findings of this study indicate that the recommended PT and AT thresholds should not be universally applied across all color measurement devices and illumination geometries due to significant inter-device variability. Instead, the results support the implementation of device-specific thresholds to improve the accuracy and consistency of instrumental shade assessment in dentistry. While ΔE_{94} and ΔE_{00} produced similar visual thresholds, the latter demonstrated superior performance in *STRESS* and *VIAS* evaluations and should therefore be preferred. However, the study further highlights the robustness and computational efficiency of ΔE^*_{ab}, which aligned with approximate perceptual uniformity within the gamut of natural tooth colors and outperformed the more complex color difference equations in this specific application.

Moreover, the x-rite MetaVue spectrophotometer was found to be well-suited for in-vitro investigations, though its design limitations restrict its applicability in clinical settings. These findings emphasize the necessity of refining current industry practices by integrating computationally validated thresholds that account for differences in device design and measurement geometries. Future research should focus on the clinical validation of these results, particularly by incorporating lay observers to assess potential commercial factor effects and by further comparing the Model-Free Estimation Technique by Zychaluk & Foster with model-based approaches such as TSK fuzzy approximation to enhance the understanding of visual color difference perception in dentistry.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests that could have appeared to influence the work reported in this paper.

Appendix

This study utilized a visual dataset that was acquired in a previous investigation, parts of which are reproduced here for comprehensive comparison. The dataset served as the foundation for computing the *STRESS* index and *VIAS* to evaluate visual-instrumental agreement, following methods detailed in a prior publication.

Table A.1
Computed ΔE_{00} color differences for each of the 26 sample pairs across all included devices. Notable differences are evident, illustrating variability in color measurement results between devices

Sample pair	eLAB	RPC	OS	PR-670	ES-V	SSM II	MetaVue
1	5.7993	5.1145	4.9738	5.2817	4.2863	5.1607	4.4357
2	5.7124	5.2935	4.3235	5.5734	4.5731	5.6640	4.2091
3	0.6769	0.5496	0.9854	0.7054	0.7495	1.5569	0.9394
4	5.1396	4.2402	4.2293	4.5712	3.3746	4.7821	4.0511
5	3.1146	3.3728	2.7620	3.2522	2.3192	2.8977	3.1929
6	1.7351	1.3821	1.3476	1.9878	1.6813	2.7046	1.6255
7	2.7908	2.5404	2.3321	2.8665	2.6416	3.4049	2.0590
8	3.0158	3.1483	2.4584	3.2110	3.0047	2.5512	2.5356
9	0.3412	0.7811	0.2306	1.2834	0.7492	1.8974	1.0598
10	2.3611	1.5842	1.8605	3.3958	1.4605	3.2340	2.5208
11	0.9264	1.4472	0.8770	2.1163	1.3038	2.5498	0.9209
12	2.0706	2.2313	2.1687	2.5665	2.5482	2.6381	2.1105
13	1.3320	1.5988	1.7634	2.8479	1.3384	2.5626	0.9962
14	3.9374	4.1600	4.0285	4.5700	4.0539	3.1043	3.8976
15	2.4020	2.4928	2.6851	2.1668	3.8640	1.3868	2.1439
16	1.8721	2.8963	2.5198	2.9715	1.5385	2.2601	1.5922
17	2.7720	2.6651	2.4745	2.9301	2.3778	1.9711	2.9894
18	1.9717	2.5081	1.7354	2.3127	2.0159	1.4302	2.2343
19	2.5571	3.2791	2.3573	2.7071	2.6551	1.7817	2.5809
20	2.4699	3.0043	2.5858	2.8215	1.4842	3.1297	1.8135
21	2.0099	1.8975	2.1025	1.2441	1.5245	1.3802	1.3133
22	3.3893	3.5745	3.1552	3.7484	2.9017	2.7722	3.0561
23	1.0965	1.0889	0.4281	0.6634	0.8815	0.9108	0.6829
24	2.7008	2.5540	2.6016	2.1080	1.9001	2.1441	2.5352
25	2.3915	2.3713	2.0574	3.4209	2.4062	2.5097	2.2542
26	1.9078	1.6357	1.4857	2.1882	1.2528	1.9537	1.3553

Table A.2
Computed color differences for ΔE_{94} per ech device and sample pair

Sample pair	eLAB	RPC	OS	PR-670	ES-V	SSM II	MetaVue
1	6.5481	5.9645	5.7114	6.1068	4.9848	6.9260	5.1224
2	5.8585	6.3294	4.9984	6.0209	5.9739	7.1127	4.1531
3	0.7046	0.6839	0.9060	0.7615	1.0324	2.0313	1.1674
4	4.4296	4.0346	3.8232	4.0309	2.9063	5.0078	3.2362
5	2.5621	2.9926	2.2156	2.7709	2.0063	2.6146	2.8320
6	1.6993	1.4016	1.2910	2.1879	2.0375	3.6456	1.7971
7	2.5382	2.4869	2.1742	2.8550	2.5469	4.4316	1.9016
8	3.1554	3.3740	2.5803	3.5663	3.3673	3.2287	2.7986
9	0.2825	0.9319	0.2630	1.5220	0.7916	2.6211	1.1998
10	2.1870	1.5812	1.7966	3.8305	1.4193	4.3322	2.3964
11	0.9590	1.3426	0.8504	2.5213	1.2872	3.5829	0.9829
12	1.9807	2.1363	1.9522	2.6734	2.4248	3.5046	2.0299
13	1.2054	1.9567	1.3843	3.6472	1.4698	3.4616	0.9305
14	3.5640	4.0352	3.5981	4.6324	3.7230	3.2485	3.5917
15	2.5820	2.6369	2.8932	2.3628	4.9527	1.4834	2.3574
16	2.0070	3.6935	2.8621	3.7006	1.7061	3.0118	1.7502
17	2.6021	2.5245	2.3293	3.1644	2.7704	2.2464	2.8035
18	1.9739	2.2252	1.6633	2.4904	2.1873	1.7678	2.0587
19	2.4038	3.1965	2.2604	2.9185	2.6490	2.1848	2.4164
20	2.8831	3.7508	2.7442	3.4671	2.0334	4.3339	2.1532
21	1.9594	2.0511	1.9189	1.2600	1.6329	1.5117	1.1965
22	3.6377	3.8006	3.4183	4.2519	3.1008	3.5091	3.3149
23	1.3258	1.4109	0.5513	0.7511	1.0608	1.2550	0.7107

(*continued on next page*)

Table A.2 (*continued*)

Sample pair	eLAB	RPC	OS	PR-670	ES-V	SSM II	MetaVue
24	2.5184	2.3995	2.3521	2.2687	2.0507	2.4608	2.3514
25	2.3875	2.4066	2.0658	3.7549	2.3792	3.3165	2.1097
26	2.1120	1.9754	1.4191	2.5850	1.2376	2.7817	1.3165

Table A.3
Computed color differences for ΔE_{00} per ech device and sample pair

Sample pair	eLAB	RPC	OS	PR-670	ES-V	SSM II	MetaVue
1	7.6983	6.8673	6.9823	6.8920	6.3592	7.1384	5.9194
2	6.3632	6.6378	5.2745	6.4280	6.2644	7.3033	4.7189
3	1.0492	0.7874	1.3910	0.8252	1.0572	2.0589	1.1942
4	5.0042	4.3428	4.2220	4.4407	3.3299	5.2352	3.6793
5	4.1193	4.7170	3.5164	4.4957	3.4724	3.2601	4.7246
6	3.1470	2.5179	2.3670	2.7723	2.8407	3.7530	2.3386
7	5.2910	4.9679	4.6610	4.8637	6.0503	4.7185	3.7342
8	5.6231	5.5893	4.6941	5.1005	5.8283	3.6509	4.0984
9	0.4434	0.9695	0.2739	1.5826	1.5630	2.6388	1.5994
10	4.5090	3.0232	3.9122	5.0819	3.3104	4.5775	4.9034
11	1.7390	1.9339	1.4942	2.7451	2.4397	3.6205	1.5207
12	4.3893	4.0212	4.1006	4.5881	5.5051	3.8544	3.9574
13	1.8209	2.1448	1.7812	3.7375	2.0075	3.5177	1.1820
14	7.5001	7.2388	6.9739	7.9086	7.8728	4.3157	7.1800
15	4.0725	3.8419	4.1152	3.4826	5.9480	1.9787	3.3807
16	2.5308	4.2391	3.5711	4.2790	2.6857	3.1723	2.6225
17	5.4094	4.5978	4.4105	4.9486	4.0646	2.8838	5.4938
18	3.1913	3.0332	2.5783	3.2597	2.9500	2.0391	2.8805
19	4.9532	5.2792	4.2793	4.4933	4.6328	2.7258	4.5968
20	3.0027	3.7974	2.9339	3.6087	2.0602	4.3776	2.2029
21	3.0677	2.7495	3.3604	2.0531	2.9730	1.6763	2.2013
22	6.0001	5.6125	5.3456	6.0873	5.4813	4.0582	5.2080
23	1.3858	1.4353	0.5523	1.0640	1.5155	1.3446	1.1585
24	5.2523	4.1533	4.5962	3.2772	3.5073	2.7950	4.6525
25	3.2821	3.3481	2.8918	5.0102	4.1586	3.5710	3.2055
26	2.8870	2.3580	2.8147	2.6510	1.9443	2.7955	1.7953

The *STRESS* index quantified the agreement between the visual shade selection and the computed selection by the devices, calculated as:

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

The VIAS score was derived directly from the *STRESS* index using the formula:

$$VIAS(\%) = 100 - STRESS$$

To compare device performance, the F-statistic was computed using the STRESS values for different devices as follows:

$$F = \frac{STRESS^2_{DeviceA}}{STRESS^2_{DeviceB}}$$

The *F*-value was then compared against a critical threshold ($F < F_C$ or $F > 1/F_C$) determined by a two-tailed *F*-distribution with a 95 % confidence interval and degrees of freedom ($N-1, N-1$), where $N = 26$ in this study. In this case, $F_C = 1.955$ and $1/F_C = 0.512$. If the *F*-value exceeded F_C or fell below $1/F_C$, the null hypothesis of no significant difference between the devices was rejected. If the *F*-value fell between these thresholds, no significant difference was assumed.

Yellow cells indicate instances where an equation performs significantly better than another, blue cells denote no significant performance difference, and grey cells signify significantly poorer performance compared to the corresponding equation in the row.

Table B.1
Individual device performance for ΔE_{00}

Device	MetaVue	ES-V	OS	eLAB	PR-670	RPC	SSM II
MetaVue	1.0	0.798	0.642	0.600	0.600	0.600	0.527
ES-V	1.308	1.0	0.823	0.769	0.769	0.769	0.676
OS	1.565	1.225	1.0	0.920	0.920	0.920	0.809
eLAB	1.666	1.303	1.048	1.0	0.979	0.979	0.860
PR-670	1.695	1.327	1.066	0.996	1.0	0.996	0.876
RPC	1.716	1.343	1.080	1.009	1.009	1.0	0.887
SSM II	1.936	1.515	1.218	1.138	1.138	1.138	1.0

Table B.2
Individual device performance for ΔE_{97}

Device	MetaVue	eLAB	ES-V	OS	PR-670	RPC	SSM II
MetaVue	1.0	0.514	0.514	0.485	0.485	0.409	0.318
eLAB	1.870	1.0	0.989	0.932	0.932	0.787	0.611
ES-V	1.903	1.007	1.0	0.948	0.948	0.801	0.621
OS	1.957	1.035	1.035	1.0	0.975	0.823	0.639
PR-670	1.979	1.047	1.047	0.986	1.0	0.833	0.646
RPC	2.362	1.250	1.250	1.177	1.177	1.0	0.771
SSM II	3.082	1.630	1.630	1.536	1.536	1.297	1.006

Table B.3
Individual device performance for ΔE^*_{ab}

Device	eLAB	MetaVue	OS	RPC	PR-670	ES-V	SSM II
eLAB	1.0	0.793	0.656	0.551	0.508	0.508	0.245
MetaVue	1.271	1.0	0.851	0.715	0.659	0.659	0.318
OS	1.452	1.176	1.0	0.817	0.753	0.753	0.363
RPC	1.724	1.397	1.154	1.0	0.894	0.894	0.431
PR-670	1.854	1.502	1.241	1.043	1.0	0.961	0.464
ES-V	1.857	1.504	1.243	1.045	0.963	1.0	0.464
SSM II	4.094	3.316	2.740	2.303	2.122	2.12	1.0

References

[1] Judd DB, Wyszecki G. Color in business, science and industry. 3rd ed. New Jersey: Wiley; 1975.
[2] 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.
[3] 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.
[4] Johnston WM. Color measurement in dentistry. J Dent 2009;37(1):e2–6. https://doi.org/10.1016/j.jdent.2009.03.011.
[5] 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.
[6] 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.
[7] Samorodnitzky-Naveh GR, Geiger SB, Levin L. Patients' satisfaction with dental esthetics. JADA 2007;138:805–8. https://doi.org/10.14219/jada.archive.2007.0269.
[8] Sghaireen Alnusayri MO, Mathew MG, Alzarea M, Bandela B, Shade V. Selection in esthetic dentistry: a review. Cureus 2022;14:e23331. https://doi.org/10.7759/cureus.23331.
[9] Berns R.S. Color and Material-Appearance Measurement. In: Billmeyer and Saltzman's Principles of Color Technology, New Jersey; 2019, p. 111- 144. http://doi.org/10.1002/9781119367314.ch6.
[10] Bazos P, Magne P. Bio-Emulation: biomimetically emulating nature utilizing a histoanatomic approach; visual synthesis. Int J Esthet Dent 2014;9:330–52.
[11] Davidson HR, Friede E. The size of acceptable color differences. J Opt Soc Am 1953;43:581–9. https://doi.org/10.1364/josa.43.000581.
[12] Choudhury AKR. Colour-difference assessment. In: Choudhury AKR, editor. Principles of colour and appearance measurement. Oxford: Woodhead Publishing; 2015. p. 55–116. https://doi.org/10.1533/9781782423881.55.
[13] Rich RM, Billmeyer FW, Howe WG. Method for deriving color-difference-perceptibility ellipses for surface-color samples. J Opt Soc Am 1975;65:956–9. https://doi.org/10.1364/JOSA.65.000956.
[14] Na N, Choi K, Lee J, Suk H-J. Color tolerance study on white in practical aspect: perceptibility versus acceptability. Color Res Appl 2014;39:582–8. https://doi.org/10.1002/col.21847.
[15] Color-Quality Specification. In: Billmeyer and Saltzman's Principles of Color Technology, New Jersey: Wiley; 2019, p. 85-110. https://doi.org/10.1002/9781119367314.ch5.
[16] Paravina RD, Ghinea R, Herrera LJ, Bona AD, Igiel C, Linninger M, et al. Color difference thresholds in dentistry. J Esthet Restor Dent 2015;27:S1–9. https://doi.org/10.1111/jerd.12149.
[17] CIE 015:2018. Colorimetry, 4th ed. Technical Report. CIE Central Bureau, Vienna. 2018.
[18] PD ISO/TR 28642:2016. Dentistry. Guidance on colour measurement. British Standards Institute; 2016.
[19] Rashid F, Farook TH, Dudley J. Digital shade matching in dentistry: a systematic review. Dent J 2023;11:250. https://doi.org/10.3390/dj11110250.
[20] Seghi RR. Effects of instrument-measuring geometry on colorimetric assessments of dental porcelains. J Dent Res 1990;69:1180–3. https://doi.org/10.1177/00220345900690051101.
[21] Alghazali N, Preston A, Moaleem M, Jarad F, Aldosari A, Smith P. The effects of different spectrophotometric modes on colour measurement of resin composite and porcelain materials. Eur J Prosthodont Restor Dent 2018;26:163–73. https://doi.org/10.1922/EJPRD_01767Alghazali11.
[22] Hampé-Kautz V, Salehi A, Senger B, Etienne O. A comparative in vivo study of new shade matching procedures. Int J Comput Dent 2020;23:317–23.
[23] 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.
[24] Kutkut N, Jordi M, Almalki A, Conejo J, Anadioti E, Blatz M. Comparison of the accuracy and reliability of instrumental shade selection devices and visual shade selection: an in vitro study. J Esthet Restor Dent 2024. https://doi.org/10.1111/jerd.13311.
[25] Hein S, Masannek M, Westland S, Spies BC, Wrbas KT, Nold J. A novel approach for the replacement of accuracy and precision measurements with the visual instrument agreement scale (VIAS) in evaluating the performance of four intraoral scanners and one shade measurement device. J Dent 2024;151:105458. https://doi.org/10.1016/j.jdent.2024.105458.
[26] Hein S, Saleh O, Li C, Nold J, Westland S. Bridging instrumental and visual perception with improved color difference equations: a multi-center study. Dent Mater 2024;40:1497–506. https://doi.org/10.1016/j.dental.2024.07.003.
[27] Gevaux L, Simonot L, Clerc R, Gerardin M, Hebert M. Evaluating edge loss in the reflectance measurement of translucent materials. Appl Opt 2020;59:8939–50. https://doi.org/10.1364/ao.403694.
[28] Bolt RA, Bosch JJ, Coops JC. Influence of window size in small-window colour measurement, particularly of teeth. Phys Med Biol 1994;39:1133–42. https://doi.org/10.1088/0031-9155/39/7/006.
[29] Lehmann KM, Devigus A, Igiel C, Weyhrauch M, Schmidtmann I, Wentaschek S, et al. Are dental color measuring devices CIE compliant? Eur J Esthet Dent 2012;7: 324–33.

[30] Ghinea R, Pérez MM, Herrera LJ, Rivas MJ, Yebra A, Paravina RD. Color difference thresholds in dental ceramics. J Dent 2010;38(2):e57–64. https://doi.org/10.1016/j.jdent.2010.07.008.

[31] Tejada-Casado M, Herrera LJ, Carrillo-Perez F, Ruiz-López J, Ghinea RI, Pérez MM. Exploring the CIEDE2000 thresholds for lightness, chroma, and hue differences in dentistry. J Dent 2024;150:105327. https://doi.org/10.1016/j.jdent.2024.105327.

[32] Manauta J, Almeida G, Kovacs-Vajna ZM, Vervack V, Shaalan O, Devoto W, et al. Precision layering techniques: integrating digital tools for accurate color matching and realistic try-ins in anterior composite restorations. J Esthet Restor Dent 2024; 36:1638–50. https://doi.org/10.1111/jerd.13297.

[33] Menini M, Rivolta L, Manauta J, Nuvina M, Kovacs-Vajna ZM, Pesce P. Dental color-matching ability: comparison between visual determination and technology. Dent J 2024;12:284. https://doi.org/10.3390/dj12090284.

[34] Zchaluk K, Foster DH. Model-free estimation of the psychometric function. Atten Percept Psychophys 2009;71:1414–25. https://doi.org/10.3758/APP.71.6.1414.

[35] Foster DH, Zchaluk K. Nonparametric estimates of biological transducer functions. IEEE Signal Process Mag 2007;24:49–58.

[36] Davidson HR, Friede E. The size of acceptable color differences. J Opt Soc Am 1953;43:581–9. https://doi.org/10.1364/JOSA.43.000581.

[37] McDonald R. Industrial Pass/Fail Colour Matching Part 1-Preparation of Visual Colour-matching Data. J Soc Dyers Colour 1980;96:372–6. https://doi.org/10.1111/j.1478-4408.1980.tb03536.x.

[38] Burns B, Shepp BE. Dimensional interactions and the structure of psychological space: the representation of hue, saturation, and brightness. Atten Percept Psychophys 1988;43:494–507. https://doi.org/10.3758/BF03207885.

[39] Meléndez-Martínez AJ, Vicario IM, Heredia FJ. Correlation between visual and instrumental colour measurements of orange juice dilutions: effect of the background. Food Qual Prefer 2005;16:471–8. https://doi.org/10.1016/j.foodqual.2004.09.003.

[40] Clarke FJJ, McDonald R, Rigg B. Modification to the JPC79 Colour–difference Formula. J Soc Dyers Colour 1984;100:128–32. https://doi.org/10.1111/j.1478-4408.1984.tb00969.x.

[41] 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;(2):103–12. https://doi.org/10.1111/jerd.12465.

[42] Ballard E, Metz MJ, Harris BT, Metz CJ, Chou J-C, Morton D, et al. Satisfaction of dental students, faculty, and patients with tooth shade-matching using a spectrophotometer. J Dent Educ 2017;81:545–53. https://doi.org/10.21815/JDE.016.022.

[43] Klinke TU, Hannak WB, Böning K, Jakstat HA, Prause E. Visual tooth color determination with different reference scales as an exercise in dental students' education. Dent J 2023;11. https://doi.org/10.3390/dj11120275.

[44] Rizzi A, Bonanomi C, Brazzoli S, Cerutti A, Kovacs-Vajna ZM. Assessing appearance in human dental colour space. Comput Method Biomec 2018;6:59–67. https://doi.org/10.1080/21681163.2016.1164079.

[45] Melgosa M, Huertas R, Berns RS. Relative significance of the terms in the CIEDE2000 and CIE94 color-difference formulas. J Opt Soc Am A Opt, Opt 2004; 21:2269–75. https://doi.org/10.1364/JOSAA.21.002269.

[46] Hein S, Morovič J, Morovič P, Saleh O, Lüchtenborg J, Westland S. How many tooth colors are there? Dent Mater 2024. https://doi.org/10.1016/j.dental.2024.10.016.

[47] Luo MR, Rigg B. Chromaticity-discrimination ellipses for surface colours. Color Res Appl 1986;11:25–42. https://doi.org/10.1002/col.5080110107.