Development of a whiteness index for dentistry

Wen Luo [a], Stephen Westland [a,*], Roger Ellwood [b], Iain Pretty [b], Vien Cheung [a]

[a] School of Design, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, UK
[b] School of Dentistry, Dental Health Unit, University of Manchester, UK

ARTICLE INFO

ABSTRACT

Objectives: The aim of the study was to develop a new whiteness index which correlates to perception of tooth whiteness.

Methods: Psychophysical experiments were conducted by a panel of observers under a standard viewing condition as well as under typical clinical viewing conditions. Various existing whiteness and yellowness indices were compared with regard to their ability to measure the perceived whiteness of human teeth. The Pearson coefficient of determination and '% wrong decision' method were used to determine the best index for tooth whiteness measurement.

Results: A new whiteness formula (WIO) was developed by optimising the original CIE whiteness formula (WIC) and it was found that WIO gave the best performance for predicting tooth whiteness based on the laboratory results. A total of 88 NHS patients repeated the same visual experiments under typical clinical viewing conditions.

Conclusion: The results confirmed the finding that the new whiteness formula is appropriate for the prediction of tooth whiteness.

1. Introduction

The color of teeth is determined by the combined effects of their intrinsic color and the presence of extrinsic stains on the tooth surface.[1] There are many treatment methods and materials available to enhance the color of teeth; these include whitening toothpaste, professional cleaning to remove stain, and bleaching agents, etc.[2,3] As tooth whitening has become a routine dental procedure, the quantification of tooth whiteness and of the efficacy of tooth whitening has been recognised and is a concern in aesthetic dentistry. Traditionally, dentists determine the color of human teeth via visual comparison to a reference standard set called a shade guide. It is a subjective process whereby the tooth and the shade guide are observed simultaneously under the same lighting conditions. Humans assessors are efficient in detecting even small differences of color between objects.[4] However, consistency between different assessors is hard to achieve. Variation in illumination, experience, age, fatigue of the human eye and color blindness may lead to inconsistencies and bias.[1] Alternatively, instrumental assessments generate quantitative and objective data. Currently, colorimeters, spectrophotometers, spectroradiometers and digital cameras are all used to measure tooth color. The measured colors are usually represented by CIE XYZ tristimulus values or CIELAB values. XYZ and CIELAB are three-dimensional color systems in which three numbers are necessary for a complete identification of any color. Whiteness is more than lightness (measured by CIE Y values or CIE L^* values); lightness, hue and colorfulness (these last two being determined by CIE a^* and b^* values in CIELAB space) of a sample all contribute towards the perception of whiteness. Whiteness is an important color attribute and there is a long history (outside of the field of dentistry) of searching for a one-dimensional color index to quantify whiteness.[5] Currently, a number of whiteness formulae are in common use including the CIE whiteness index WIC, the whiteness index according to ASTM E-313-73 WI, and the Z% index.

Since the color of most human teeth corresponds to a small range of the color space from yellowish white to light brown and the degradation and aging of teeth is usually associated with yellowness, some researchers have proposed that changes in yellowness are important factors in the assessment of tooth whitening.[6] As a consequence, the quantification of whiteness has sometimes been expressed by a quantification of yellowness; lower yellowness values are considered as corresponding to higher whiteness. In a study to assess whether whiteness or yellowness indices can predict the perception of tooth whiteness WIC gave the best result for measuring tooth whiteness when compared with visual assessments.[7] However, the applicability of such formulae in dentistry is uncertain since many were designated to be used only with samples whose color coordinates are within a narrow range and the colors of teeth appear to be outside this range.[8] There is a clear need for psychophysical studies of tooth-whiteness perception in order to validate the performance of various metrics and to develop a proper index for tooth whiteness. This paper reports such a study and develops a whiteness index with specific intent for use in dentistry.

2. Materials and methods

Psychophysical experiments were conducted where observers made judgements on the whiteness of tooth-color samples under standard laboratory viewing conditions and, separately, under less rigorous clinical viewing conditions.

2.1. Color measurements

Two sets of Vitapan shade samples (the Toothguide 3D Master shade guide and the Vitapan Classical shade guide) were used in the visual experiments. The Toothguide 3D Master shade guide consists of five groups of tabs according to their lightness level where all the tabs in one group are designed to have the same lightness. The tabs go from lightest to darkest when moving from left to right. Within each group there are only differences in chroma and hue. There are three

kinds of hue from yellowish to reddish for Groups 2, 3 and 4, whereas Groups 1 and 5 have only one hue. The chroma of tabs decreases from top to bottom in the arrangement.[9] The Vitapan Classical shade guide consists of 16 tabs with a simpler arrangement and can be arranged from lightest to darkest.[10] Neither shade guide presents a perceptually linear whiteness scale because the perceptual color difference between adjacent tabs is not a constant.

The use of a tele-spectroradiometer TSR to assess tooth color is a non-contact and objective method that has been used in dentistry.[11–13] A Minolta CS1000 TSR was used in this study to measure the spectral reflectance factors (and hence the color coordinates) of the shade guide tabs since it can measure colors in a way that matches the geometry of the visual assessments (samples positioned about 40 cm away from the subject and viewed at an angle of about 45°). The setup for measurement using the Minolta CS1000 is shown in Fig. 1.

A VeriVide CAC60 viewing cabinet (CIE illuminant D65) was employed to provide consistent viewing conditions. Since teeth are translucent and the oral cavity behind teeth is relatively dark during typical conditions under which teeth are viewed, a black background was used for measurements in this study. A triangular stand was built to hold tooth samples to avoid the specular reflection from the glossy surface plots the measured a^* and b^* values of the Vitapan Toothguide 3D Master shade guide tabs and the Vitapan Classical shade guide. As the 3D-shade guide has a better arrangement and coverage of tooth color, it was chosen as the training set for developing the tooth whiteness index, whereas the Vitapan Classical shade guide was applied as a testing set to validate the developed whiteness index in both standard and clinical viewing conditions (Fig. 2).

2.2. Psychophysical experiments

Two psychophysical experiments were conducted by two groups of observers. In the first experiment (laboratory), observers were asked to rank the tabs from each of the shade guides under a controlled viewing condition. Results from this
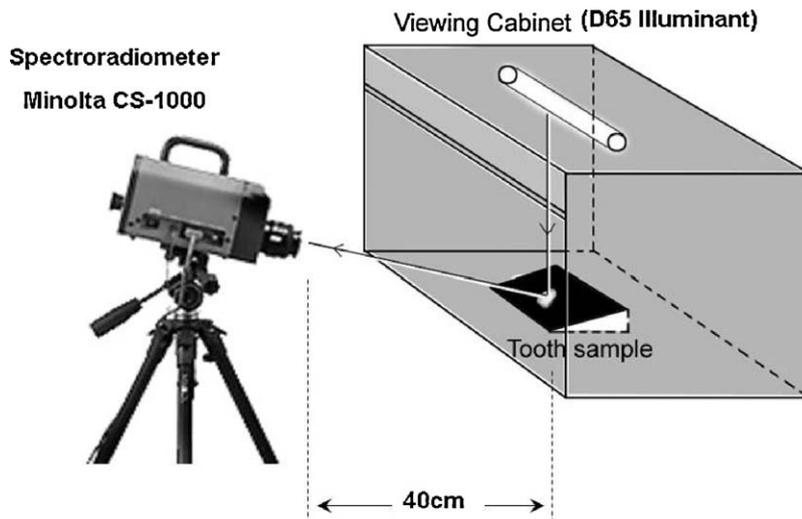

Fig. 1 – The setup for measuring teeth samples using the Minolta CS1000 spectroradiometer.
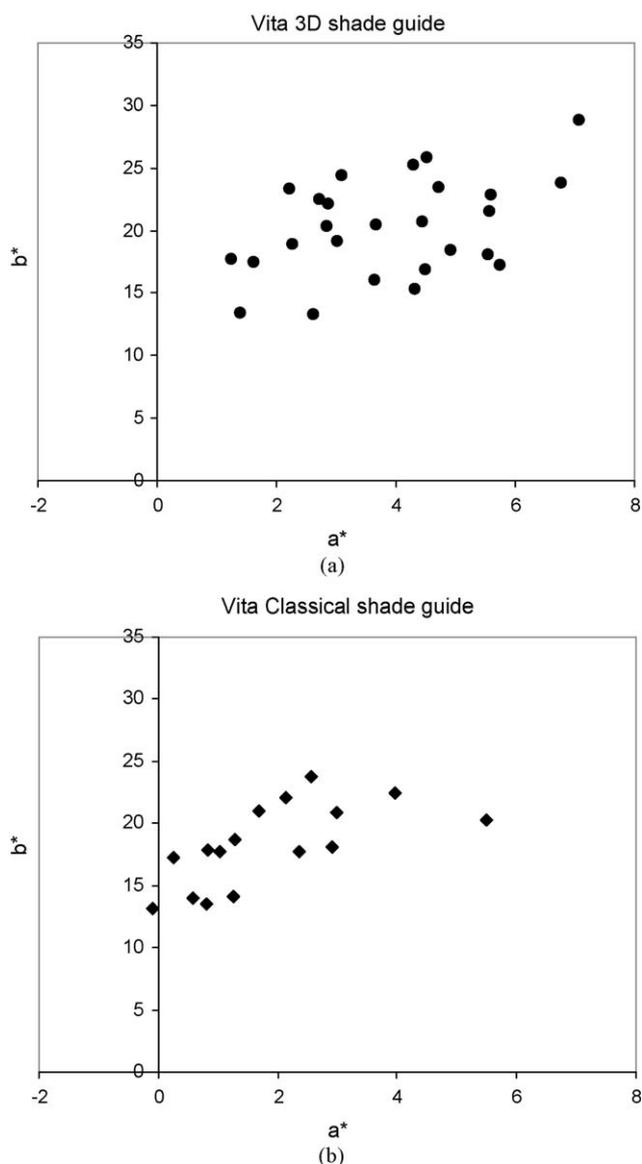
Fig. 2 – Measured a^* and b^* values of (a) the Toothguide 3D Master shade guide and (b) the Vitapan Classical shade guide.

experiment were used as the training data for developing the whiteness index for teeth. The second experiment (clinical) was to ask a larger group of observers to repeat the visual ranking judgements under typical clinical viewing conditions.

In the laboratory experiments 9 observers (aged 22–38, 4 males and 5 females), all of whom passed the Ishihara color vision test,[14] visually ranked the tabs according to whiteness. Observers sat approximately 40 cm from the centre of the viewing cabinet; the viewing angle was approximately 45° to avoid the specular reflection. Before ranking the tabs, observers adapted to the grey interior of the viewing cabinet for approximately 60 s.[15] There was no time restriction for observers; they continued until they were satisfied with the rank order. Tabs from the Toothguide 3D Master and Vitapan Classical shade guides were separately ranked. For ease of reference, the two experiments were notated as *Vita-3D Experiment* and *Vita-Classical Experiment*. The *Vita-3D Experiment*

was repeated after three days to check the repeatability error of observers.

In the clinical experiment 88 of observers (drawn from NHS dental patients) were asked to rank the 16 tabs from the Vitapan Classical shade guide in order of perceived whiteness under typical clinical lighting conditions. The purpose of the *Clinical-Vita-Classical Experiment* was to investigate whether the WIO index would give good agreement with visual observations under typical clinical (non-laboratory) conditions made by naive observers using the testing set of tabs.

2.3. Whiteness metrics

The initial candidate whiteness metrics considered in this study were three widely used whiteness indices (WIC, Z%, WIE313), the CIELAB b^* value (b^*), and two yellowness indices (E313, D1925).

The CIE whiteness formula WIC was recommended in 1986[5,16,17] and with neutral hue preference is:

$$WIC = Y + 800(x_n - x) + 1700(y_n - y) \tag{1}$$

where (x, y) and (x_n, y_n) are the chromaticity coordinates of the sample and the reference white respectively. The Z% whiteness index is computed from the CIE Z value, where Z_n is the CIE Z value of the reference white, thus,

$$Z\% = \frac{100Z}{Z_n} \tag{2}$$

The ASTM E313 whiteness index WIE313 that combines Z% with luminance factor was evaluated thus,

$$WIE313 = 4Z\% - 3Y \tag{3}$$

The ASTM E313 and D1925 yellowness indices were also evaluated.

$$YIE313 = 100 \left(1 - \frac{0.847Z}{Y}\right) \tag{4}$$

$$YID1925 = \frac{100(1.275X - 1.057Z)}{Y} \tag{5}$$

The whiteness/yellowness indices of each tab from the shade guides were calculated from the CIE XYZ values measured by the TSR.

An optimised version of the CIE Whiteness Index WIO was introduced. The original linear form suggested by Ganz[16] was retained but the coefficients of the equation were optimised to achieve the best fit between the z scores from the visual assessments and the values from the index. The coefficients of the WIO formula were determined by achieving the maximum correlation of determinations to the visual results through TSR nonlinear optimization[21] provided by the Microsoft Excel Solver tool. This led to the following equation:

$$WIO = Y + 1075.012(x_n - x) + 145.516(y_n - y) \tag{6}$$

Table 1 – Evaluation results of the existing whiteness metrics against the visual result of the Toothguide 3D Master shade guide.							
	Obs.	WIO	WIC	Z%	YID1925	YIE313	b^*	WIE313
r^2	–	0.93	0.87	0.85	0.85	0.81	0.62	0.56
WD (325 pairs)	5.71	5.80	13.54	8.92	12.31	16.00	22.46	25.54
WD (25 pairs)	8.44	8.00	16.00	16.00	24.00	28.00	32.00	32.00

2.4. Visual ranking data

The simplest method of data analysis would be to compare the ranking orders from observers with the whiteness index values. However, raw ranking data allow comparisons of the type "whiter than" or "darker than," but do not reveal the magnitude of the differences. A more sophisticated statistical scaling technique is preferred. This scaling procedure, described by Engen,[18] applies the law of comparative judgements to transform the ordinal scale ranking data to an equal-interval scale. It converts rank orders to z scores (in units of normal deviates). The z scores obtained represent values for the stimuli on a psychological scale with equal intervals, which provide an indication of relative differences amongst the stimuli but which are, nevertheless, not meaningful in absolute terms.

2.5. Performance metrics

Two performance metrics, coefficient of determination r^2 and "% wrong decisions" WD, were used to aid the comparison between the visual interval-scale data and the whiteness values. The linear relationship between the two data sets was tested by the coefficient of determination. However, in order to infer whether the use of any of the whiteness formulae would be adequate in practical terms, an additional analysis based upon the statistical method called '% wrong decisions' was applied.[19,20] A wrong decision is defined as one where an individual observer's judgement disagrees with the majority view. The results were presented as WD, the percentage of times that an observer made a wrong decision. Observer accuracy was quantified by calculating values of the coefficient of determination (r^2) between the data for each individual observer and the mean of the panel results. Observer repeatability was tested by comparing the results of the two repeat sessions in the Vita-Classical Experiment.

3. Results

3.1. Observer variability

The results of observer accuracy for the Vita-3D Experiment, the Vita-Classical Experiments and the Clinical-Vita-Classical Experiment were 0.93, 0.94 and 0.92, respectively. The agreement between the two sessions was 0.97.

3.2. Performance of metrics

For the laboratory Vita-3D Experiment the correlation coefficients are listed in the first data row in Table 1. WIC has the best performance of the various published indices with r^2 of 0.87, whereas the b^* and WIE313 correlate poorly with the visual results.

It was found that the correlation coefficients of most these formulae were quite high, but it is difficult to infer whether the use of any of the whiteness formulae would be adequate in practical terms. The additional WD analysis was used in this study to help to address this question. When the 325 possible paired comparisons for the 26 tabs from the Toothguide 3D Master shade guide are considered the mean wrong-decision performance of the observers was found to be 5.7%. The same method was applied to the indices (comparing the instrumental decision with the visual consensus) and the percent-wrong-decisions are listed in the second data row in Table 1. When the analysis was repeated with just 25 paired comparisons (where each pair now consisted of the adjacent samples in the average visual ranking) the mean observer percent-wrong-decisions were 8.4%. Again the performance of all of the whiteness metrics was substantially worse than the observer performance.

The correlation performance of the optimised equation WIO was 0.93. WIO yields only 5.8% wrong decisions for the 325 pairs and 8% wrong decisions for the 25 pairs.

3.3. Clinical evaluation of the new whiteness index

The coefficient of determination r^2 between the Vita-Classical Experiment and the Clinical-Vita-Classical Experiment was 0.98. Thus observers can be considered to give nearly identical ranking results even under different viewing conditions for the same teeth sample set.

Table 2 summarises the coefficient of determination between the visual rankings (expressed in z scores) pooled for the 9 and 88 observers and the various metrics for the 16 tabs of the Vitapan Classical shade guide. It is noted that even though the WIO formula was optimised using the visual assessments of 9 observers for the 26 tabs of the Toothguide 3D Master shade guide, the WIO formula gives good agreement with the new data based on 88 observations of the 16 samples of the Vitapan Classical shade guide made under typical clinical conditions and also for the 9 observers who ranked these samples under laboratory conditions. WIO gave the highest value of the coefficient of determination, followed by WIC index, whereas the yellowness indices, especially b^*, correlated poorly to the visual results.

The average percent-wrong-decisions for the 88 observers in the clinical experiment were 6.20 and 16.44 for the 120- and 15-pair comparisons. For comparison the equivalent scores for the 9 observers in the laboratory experiment were 4.63 and 17.04, respectively. Fig. 3 summarises the WD results of the Vitapan Classical shade guide for the 120 and 15 paired comparisons in more intuitive version. It shows that the predictions of tooth whiteness using the previously published

whiteness and yellowness formula may not be acceptable since they are less reliable than visual assessments made by the average human observer.

4. Discussion

Since different observers may have different hue preference in whiteness assessment, it might be easier to repeat the same orders of ranking by the same observer than to achieve agreement between different observers. It is evident that if any of the indices were used for predicting whiteness, the performance would be worse on average than that of a human observer selected at random (for example, the use of b^* would result in four times as many wrong decisions as the average observer). We argue that although WIC strongly correlates with the visual assessments ($r^2 = 0.87$) it still cannot be considered to be adequate since its use would generate about 13.5% wrong decisions (compared with a mean visual performance of 5.7%). It was found (Table 1) that the performance of all indices was worse than that of the average observer with the exception of the WIO index whose performance alone can be considered to be adequate. Note that both WIC and WIO formulae give relative, but not absolute, evaluations of whiteness. The difference just perceptible to an experienced visual assessor is about 3 CIE whiteness units.[22] The formulae were devised to offset the lightness of a sample with its color. Therefore, the lighter a tooth, the whiter it is. But for two equally light teeth the one that is yellower will have a smaller WIO value. The measurements by Garcia[7] showed that the WIC values for 20 teeth samples were in a range of −23.5 to −118.1. In this study, the WIO values of the tabs from Toothguide 3D Master shade guide were from −0.6 to −81.5.

In the clinical study only the WIO formula gives performance that is comparable to that of the average visual performance and we therefore conclude that only this formula can be said to be acceptable in the context of practical use. The results in Table 2 and Table 3 all indicate worse performance than the first visual ranking study during which a smaller number of observers made visual judgments under carefully controlled conditions. However, the relative performance of the metrics was similar for the two studies.
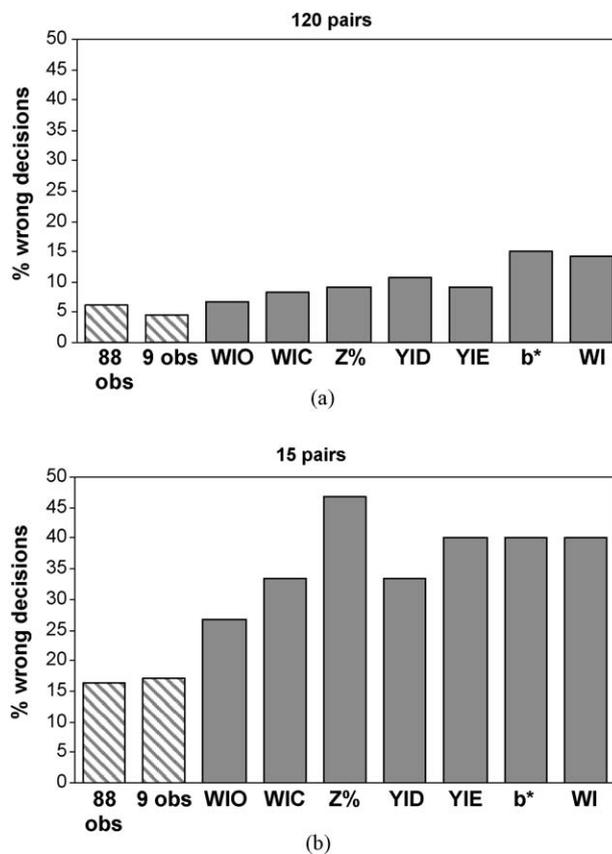

Fig. 3 – The percent-wrong-decisions for observers and various indices of the Vitapan Classical shade guide (a) 120 pairs; (b) 15 adjacent pairs.

5. Conclusions

Various whiteness and yellowness indices were compared for their ability to predict the perception of tooth whiteness. According to visual results, it was found that the CIE whiteness formula WIC has the best performance amongst the existing published formulae that were tested. However, a further analysis showed that the predictions of tooth whiteness using the WIC formula are less reliable than visual assessments made by the average human observer. A modified version of the

Table 2 – Pearson coefficient of determination between the various indices and z scores from visual assessments by 9 and 88 observers.							
r^2	WIO	WIC	Z%	YIE313	YID1925	WIE313	b^*
9 Obs.	0.92	0.87	0.82	0.81	0.74	0.66	0.66
88 Obs.	0.85	0.79	0.79	0.72	0.78	0.59	0.57

Table 3 – Percent-wrong-decisions for observers and the various indices based on the full set of 120 paired comparisons and the 15 color-critical paired comparisons of adjacent samples.								
% error	88 Obs.	9 Obs.	WIO	WIC	Z%	YID	YIE	b^*	WI
120 pairs	6.20	4.63	6.67	8.33	9.17	10.83	9.17	15	14.17
15 pairs	16.44	17.04	26.67	33.33	46.67	33.33	40.00	40.00	40.00

whiteness formula WIO was developed by optimisation of the coefficients of the WIC according to the visual results for the Vita 3D-shade guide. It was found that WIO outperformed the WIC formula and was further shown to be acceptable in that it was as reliable as the average human observer based on the Vita 3D-shade guide laboratory data. Psychophysical studies involving 88 observers under typical clinical viewing conditions were carried out and the data from these studies were best predicted using WIO. Although the WIO index was optimised using observations made under laboratory conditions of a different set of porcelain teeth samples, it gave good performance when used to predict perceptual whiteness.

Conflict of interest statement

Acknowledgement

REFERENCES

1. Watts A, Addy M. Tooth discoloration and staining: a review of the literature. *British Dental Journal* 2001;**190**:309–16.
2. Joiner A. Tooth color: a review of the literature. *Journal of Dentistry* 2004;**32**:3–12.
3. Joiner A, Hopkinson I, Deng Y, Westland S. A review of tooth color and whiteness. *Journal of Dentistry* 2008;**36S**:S2–S7.
4. Paul S, Peter A, Pietroban N, Hammerle CHF. Visual and spectrophotometric shade analysis of human teeth. *Journal of Dental Research* 2002;**81**:578–82.
5. Ganz E. Whiteness measurement. *Journal of Color and Appearance* 1972;**1**:33–41.
6. Barlow AP, Brennan K, Desai AP. Clinical comparison of two self-directed at-home bleaching systems. *Proceeding of the 81st general session of the international association for dental research.* 2003:900.
7. Garcia JA, Rubino M, Romero J, Hita E. Measuring the whiteness of human teeth. *Journal of Color Research and Application* 1993;**18**:349–52.
8. Rubino M, Garcia JA, Barco L, Jimenez del. Romero J. Color measurement of human teeth and evaluation of a color guide. *Journal of Color Research and Application* 1994;**19**:19–22.
9. Browning WD. Use of shade guide for color measurement in tooth-bleaching studies. *Journal of Esthetic and Restorative Dentistry* 2003;**15**:S13–20.
10. Buisson JC, Baker RA, Carpinello LA, Gerlach RW. Relevance of standard shade tab ranking to personal color preference. *Proceedings of 80th general session of the international association for dental research.* 2002:27.
11. Hasegawa A, Motonomi A, Ikeda I, Kawaguchi S. Color of natural tooth crown in Japanese people. *Journal of Color Research and Application* 2000;**25**:43–8.
12. Reno EA, Poore CL, Anastasia MK, Miller JM, Crisanti MM. Reproducibility of non-contact tooth color measurement system. *Proceeding of 30th annual meeting of the AADR.* 2001:1365.
13. Martin-delasHeras S, Valenzuela A, Bellini R, Salas C, Rubino M, Garcia JA. Objective measurement of dental color for age estimation by spectroradiometry. *Journal of Forensic Science International* 2003;**132**:57–62.
14. Ishihara S. Test for color-blindness, 16 plates edition. Tokyo: Kanehara Shuppan Co., Ltd.; 1962.
15. Fairchild MD, Reniff L. Time course of chromatic adaptation for color appearance judgements. *Journal of the Optical Society of America* 1995;**A12**:824–33.
16. Ganz E. Whiteness: photometric specification and colorimetric evaluation. *Journal of Applied Optics* 1976;**15**:2039–58.
17. Ganz E. Whiteness formula: a selection. *Journal of Applied Optics* 1979;**18**:1073–8.
18. Engen T. Psychophysics: scaling methods. In: Kling JW, Riggs LA, editors. *Woodworth and Schlosberg's experimental psychology.* 3rd ed. New York: Holt, Rinehart, & Winston, Inc.; 1971.
19. McLaren K. Color passing—visual or instrumental. *Journal of Society of Dyers and Colorists* 1970;**86**:389–93.
20. MacDonald R. Industrial pass/fail color matching. Part I. Preparation of visual color-matching data. *Journal of Society of Dyers and Colorists* 1980;**96**:372–6.
21. Lasdon LS, Waren AD, Jain A, Ratner M. Design and testing of a generalized reduced gradient code for nonlinear programming. *ACM Transactions on Mathematical Software* 1978;**4**:34–49.
22. Hayhurst R, Smith K. Instrumental evaluation of whiteness. *Journal of the Society of Dyers and Colorists* 1995;**III**:263–6.