Neuro-Ophthalmology & Visual Neuroscience, 2018, Vol. 3, No. 1, 1-6
Available online at http://pubs.sciepub.com/novn/3/1/1

DOI:10.12691/novn-3-1-1

The Repeatability of Colorimetry is Precise(ly) as Expected

Amelia Aldrich[1], Rupal Lovell-Patel[2,3], Peter Allen[2,3], Arnold Wilkins[1,*]

[1]Department of Psychology, University of Essex, Colchester UK
[2]Department of Vision and Hearing Sciences, Anglia Ruskin University, Cambridge UK
[3]Vision and Eye Research Unit, Anglia Ruskin University, Cambridge UK
*Corresponding author: arnold@essex.ac.uk

Abstract The purpose of the study was to assess the repeatability of clinical assessments with the Intuitive Colorimeter, a repeatability classified as "poor" in a previous study. Patients underwent assessments with the Intuitive Colorimeter in two studies. In each study, one published by Suttle et al [1] and the other described herein, assessments were undertaken on two occasions. The studies differ in respect of the models of colorimeter used, the methods employed, the interval between examinations, and the masking of examiners. The repeatability was assessed using the methods currently used in clinical practice, which differ according to examiner. Both studies show a similar repeatability of the assessments. This repeatability is consistent with previous literature. We estimate the standard deviation of u' and v' coordinates each to be 0.020 and thereby obtain an estimate of the number of tinted trial lenses necessary when prescribing coloured filters. In patients with visual stress assessment with the Intuitive Colorimeter is repeatable. The minimum number of tints necessary for assessment is estimated to be 77.

Keywords: *Intuitive Colorimeter, precision tints, dyslexia, migraine, reliability*

Cite This Article: Amelia Aldrich, Rupal Lovell-Patel, Peter Allen, and Arnold Wilkins, "The Repeatability of Colorimetry is Precise(ly) as Expected." *Neuro-Ophthalmology & Visual Neuroscience*, vol. 3, no. 1 (2018): 1-6. doi: 10.12691/novn-3-1-1.

1. Introduction

A technical note in Ophthalmic and Physiological Optics has suggested that the repeatability in choosing coloured overlays and lenses is "poor" [1]. It is shown here that the repeatability of colorimetry is consistent with previous literature, and a study is reported that confirms the repeatability under masked conditions.

The Intuitive Colorimeter is an instrument that is used to obtain a tint that reduces visual stress. It illuminates text with coloured light and permits the separate manipulation of hue and saturation without an associated change in luminance. In the first such instrument [3], light from three coloured filters was mixed in different proportions. In a subsequent design [4], seven coloured filters were used to ensure that the spectral power distribution of the light in the instrument closely resembled that obtained when coloured spectacle lenses were worn under typical office lighting.

The process of tint selection is subjective. The text is illuminated by light of one hue and the saturation increased and decreased at that hue in order to assess all shades of this hue in comparison with white. The hue is then changed and the process repeated for 12 hues. Having short-listed those of the 12 associated with an improvement in visual comfort, the hue and saturation are alternately adjusted by small amounts so as to evaluate hues that lie between those originally examined. The best overall combination of hue and saturation is thus obtained under conditions in which the eyes remain colour adapted. Combinations of tinted trial lenses are then offered that match the chosen chromaticity under office lighting. From seven dyes are chosen two with neighbouring chromaticity. Each dye has five lenses with a geometric progression of dye deposition. By combining the lenses 2^5 -1 = 31 levels of deposition are available for each dye. When the two dyes are combined the 31*32=992 combinations of lenses sample the appropriate chromaticities so as to provide a visible match to any colorimeter setting.

The Intuitive Colorimeter described above was initially designed as a research tool to investigate the improvement in visual comfort with different colours [5] but has subsequently found clinical use in eye-care practice.

Suttle et al classify colorimetry as having "poor" repeatability because the number of Just Noticeable Differences (JNDs) between two measurements is large. JND refers here to the difference that is just noticeable when two coloured surfaces are observed side by side, as is the case with coloured overlays. JNDs have previously been used [2] as an adjunct to the unfamiliar concept of chromaticity difference, with the McAdam ellipses as a guide. When two surfaces are simultaneously visible, very small differences in chromaticity are discernible, partly as a result of opponent processes. It is not clear from the paper by Suttle et al how the JNDs were derived, but it could be argued that the use of JNDs is inappropriate in the context of colorimetry. During colorimetry coloured lights are presented successively. Any comparison involves

not only adaptation to the colour but memory for colours previously shown. This inevitably increases variability. JND is a subjective measure, whereas chromaticity difference is a physical measure that is directly related to the relative energy captured by the three classes of cones. For this reason only chromaticity difference is used in this paper. It is shown that the repeatability of assessment with the Intuitive Colorimeter is as would be anticipated from previous literature.

Some idea of the likely repeatability of the colorimetry assessment can be obtained in various ways, as will now be described.

The first method of estimating the likely repeatability is from studies that presented text under coloured light and repeatedly required observers to report any reduction of perceptual distortion, recording the chromaticities associated with such reduction. The paper that first described the device also described such a study [6]. A small sample of three children reported a reduction of perceptual distortion with light of particular chromaticities. From the graphs provided for each individual, the average separation between all pairwise combinations of chromaticities has been obtained, and the average (SD) was 0.046 (0.026) for participant A, 0.064 (0.03) for participant B and 0.052 (0.032) for participant CD (a participant who was examined on two occasions, C and D).

A second method of judging the likely repeatability involves describing the boundary in the chromaticity diagram within which perceptual distortions abate. This method was used in a subsequent paper [2] that reported a double-masked study. Twenty-three participants were asked to select a chromaticity that reduced their perceptual distortions. The hue was then altered progressively until the distortions returned. The average difference in chromaticity between the two settings was 0.065 [2].

A third method of judging the likely repeatability is more objective and involves measuring the effect of the tint on reading speed. Five individuals who used coloured lenses were asked to read randomly ordered common words aloud rapidly, without their lenses, under light of randomly chosen chromaticities [7]. The reading speed was plotted as a function of the chromaticity difference between the colour used for reading and that selected as optimal for clear perception of the text. It can be seen from Figure 1 and Figure 2 in reference [7], which show the variability of the observations, and from Table 1 in reference [7], which shows the parameters of the curve fitting, that for all individuals, there was a progressive decrease in reading speed with chromaticity difference. Little beneficial effect of the colour was apparent when the chromaticity difference reached 0.07 (ΔE^* = 91, see Appendix of reference [7]).

From the above studies, involving 3, 23 and 5 individuals respectively, it appears that the beneficial effects of the colour are lost when the chromaticity of a light differs by 0.07 from that selected as optimal for the perception of the text as "clear and comfortable". This inference is supported in four studies involving more than 80 participants in which active and placebo lenses differed in chromaticity by 0.065 - 0.07. Two were masked studies of symptoms that improved with active lenses more than with the control [2,8]. Two involved measurement of the characteristics of the cortical haemodynamic response to uncomfortable patterns, and showed a normalization only with the active tint and not the control [9,10]. In all four studies, both those with and without high rates of attrition, the effects of the optimal tint were greater than those of the placebo. This indicates that a chromaticity difference of 0.07 is sufficient to reduce the clinical efficacy, and sets limits on the effective chromaticity.

In the paper by Suttle et al. [1] participants were considered to have visual stress if they reported one or more symptoms of visual discomfort or distortion while reading, and they reported alleviation of symptom(s) with a coloured overlay. This selection process is possibly more lenient than implied by the practical diagnostic guidelines suggested recently [11] but the study was conducted before these were available. Participants were asked to view text and to report any symptoms experienced. If any symptoms were reported, participants were then asked to observe the text through each of the ten Intuitive Overlays (starting with Rose), individually and then in combinations of two, following the procedure previously described [12].

This was followed by assessment with the Intuitive Colorimeter [13] in a darkened room, following the procedure described earlier. The assessment was repeated at a second appointment between 2 and 57 days after the first. Only four individuals in the sample of 20 showed a difference in colorimeter measurements that was greater than 0.07. All four had poor consistency in the choice of overlay.

Numerical methods were used to estimate the probability of chance occurrence of colorimeter settings as close as those obtained. The first settings were re-paired with the second settings randomly across participants, and the average chromaticity difference calculated. This was done 1000 times and 34 such re-pairings gave a chromaticity difference less than that obtained from the correct pairing. On this basis the chance occurrence of the chromaticity difference obtained may be estimated to be 0.034.

Given the above, the repeatability of colorimetry judgments was as might have been expected from the previous literature. It was also somewhat better than might have been expected had observers simply memorised the hue. D'Ath et al [14] required healthy observers to reproduce a previously displayed screen colour in a darkened room by varying hue (h_{uv}) with saturation constrained to lie on a circle radius 0.060, centred on the chromaticity of D65. The average difference between the original hue (h_{uv}) and that reproduced one hour later was 48 degrees; a difference in chromaticity of 0.049. This difference may be compared with the difference of 0.035 in the previous paper [1] with the "consistent" participants (those who chose overlays of the same or similar colour).

In the following study colorimetry was undertaken twice on the same day by two independent masked examiners using different instruments and the procedures in current clinical practice, so as to assess the repeatability in typical use.

2. Methods

2.1. Participants

Twenty participants were recruited from patients attending the Anglia Ruskin University Eye Clinic for

assessment of visual stress. (See Table 1). Written consent was obtained from the individual or a parent. All methods were approved by the Anglia Ruskin University Faculty of Science and Technology Research Ethics Panel. The details of the patients and their results are given in Table 1. The extent to which the patient selection conformed to the criteria for visual stress recently introduced by Evans et al [11] can be judged from Table 1. 15 participants had used an overlay for >3 months. All read more quickly with an overlay but only in 16/20 was the increase in reading speed greater than 15%; 15/20 had Pattern Glare scores of more than 3; 17/20 had 3 or more symptoms. Overall, 16 patients satisfied the criteria [11].

Table 1. Patients' details, reason for referral, Delphi criteria for visual stress (VS), consistency ratings, colorimetry results and rate of reading. The signs of visual stress (Table 4 of Reference [11]) are shown as present (1) or absent (0) in columns 4-6, and the number of symptoms is shown in column 7

Sex	Age	Reason for Referral	3 month overlay use	Increase in RRT >15%	Pattern Glare >3	Symptoms from text	Pass Delphi criteria	1st Reliability Rating	2nd Reliability Rating	1st Colorimetry	2nd Colorimetry	Difference in hue (degrees)	1st Colorimetry		2nd Colorimetry		Difference in chromaticity	Reading rate Active lens	Reading rate Placebo lens
										hue, satn	hue, satn		u'	v'	u'	v'		(wpm)	(wpm)
M	8	VS/Dyslexia	0	0	0	1	0	Good	Good	300 30	300 50	0	0.252	0.473	0.263	0.449	0.026	103	96
M	12	VS/ Concussion	1	1	1	3	1	Good	Good	120 30	150 50	30	0.215	0.543	0.156	0.553	0.059	188	184
M	14	VS/Dyslexia	1	1	0	3	1	Good	Good	180 30	260 40	80	0.179	0.528	0.179	0.452	0.076	119	78
F	17	Dyslexia	1	1	1	3	1	Good	Good	165 30	170 35	5	0.180	0.534	0.188	0.535	0.008	217	216
F	17	Headaches/ Dyslexia	0	1	1	3	1	Good	Good	80 30	80 40	0	0.253	0.541	0.240	0.554	0.018	119	117
M	18	VS	1	1	1	3	1	Good	Good	300 30	330 50	30	0.253	0.474	0.297	0.479	0.044	123	131
M	19	VS/Photo-sensitivity	1	1	0	3	1	Good	Good	150 30	180 40	30	0.182	0.537	0.169	0.531	0.015	91	86
F	23	VS	1	1	1	6	1	Good	Good	180 35	150 40	30	0.164	0.528	0.173	0.547	0.021	164	182
F	27	VS	1	1	1	3	1	Good	Good	30 30	30 40	0	0.286	0.530	0.278	0.539	0.012	166	159
M	31	VS/Dyslexia	1	1	1	3	1	Good	Good	270 30	270 30	0	0.270	0.463	0.207	0.485	0.067	100	75
F	38	VS/Dyslexia	1	1	1	3	1	Good	Good	180 30	210 50	30	0.179	0.527	0.146	0.508	0.038	150	124
F	9	VS	1	1	1	3	1	Good	Moderate	90 30	70 50	20	0.238	0.544	0.255	0.559	0.023	70	85
F	11	Dyslexia	1	1	1	3	1	Good	Moderate	30 30	30 30	0	0.287	0.530	0.252	0.535	0.036	77	51
F	13	VS	0	1	1	4	1	Good	Moderate	150 30	150 50	0	0.182	0.538	0.157	0.553	0.029	136	114
F	23	Dyslexia	1	1	1	4	1	Good	Moderate	130 30	60 30	70	0.203	0.541	0.242	0.540	0.040	276	271
M	71	VS	1	1	1	3	1	Moderate	Good	0 30	300 50	30	0.300	0.518	0.262	0.450	0.078	115	104
M	9	VS/Dyslexia	1	1	0	3	1	Poor	Good	150 30	150 40	0	0.181	0.537	0.169	0.549	0.017	93	84
F	13	Headaches	1	0	1	0	0	Poor	Poor	180 25	10 35	170	0.191	0.527	0.283	0.528	0.092	102	95
M	14	VS/Dyslexia	0	0	0	0	0	Poor	Poor	290 30	260 50	30	0.238	0.470	0.165	0.424	0.086	79	74
M	58	VS	0	0	1	5	0	Poor	Poor	150 20	330 35	180	0.203	0.531	0.267	0.499	0.071	148	144

2.2. Procedure

Participants underwent an extended eye examination and colorimeter assessment that was conducted by optometry students in the clinic under the supervision of RLP. An Intuitive Colorimeter Mk. 3 (Cerium Visual Technologies, Tenterden, Kent) was used to identify a colour of illumination that maximized comfort and reduced perceptual abnormalities.

The colorimetry assessment followed the practice at Anglia Ruskin of restricting the saturation to 30. The assessment is thereby simplified for students, and avoids the provision of hues that are too dark; the final assessment is reviewed by the supervisor, and revised if necessary. The second assessment did not restrict the saturation and followed the procedure described in the Colorimeter Manual (Cerium Visual Technologies) [13]. The difference in the colorimetry procedure will have increased the difference in the chromaticities obtained, but the central concern in this paper was to study the repeatability of colorimetry under the conditions that currently obtain in clinical practice, including differences in procedures and instruments.

Two lens stacks were prepared using the trial lenses supplied with the colorimeter and worn in a lens holder mounted on a headband. One stack was *active* (matching under office lighting the chromaticity selected by the participants as optimal) and one *placebo* having CIE 1976 UCS chromaticity that differed by an average of 0.078 (SD = 0.02). The placebo was selected by a spreadsheet that chose a similar combination of lenses but with different dyes. There are seven dyes in the colorimeter system that can be arranged in a hue circle. The placebo lenses differed by two steps on this circle and had a transmission similar to that of the active lenses.

A second examiner (AA) who was not party to the identity of the two lens stacks administered the Wilkins Rate of Reading Test four times, A-D. One set of lenses, chosen at random, was used for passages A and D and the second set for B and C.

The patient was then debriefed to allow them to leave if they wished. All consented to remain and underwent a second colorimeter assessment conducted about one hour after the first by examiner AA. (In normal observers, memory for coloured lights after one hour is similar to that

after one week [14]). The second assessment used an Intuitive Colorimeter Mk. 2. This machine has the same filters as the Mark 3, but has a very different appearance. The second examiner had no knowledge of the results of the first examination.

Both the examiners who undertook the colorimeter examination gave a rating reflecting their assessment of the certainty with which patients were able to choose their optimal chromaticity. Individuals who identified an optimal colour and did not deviate, were given a consistency rating of "good"; those who were unsure or for whom the colour was repeatedly adjusted were rated "poor." Those participants who identified a colour with only minimal adjustment were rated "moderate." The examiners gave their ratings independently without knowledge of the rating given by the other examiner.

3. Results

Participants as a group read more quickly with the 'active' set of lenses: 132 words per minute compared to 123 words per minute with 'placebo', one-tailed $t(19) = 2.61$, $p = 0.0086$, $d=0.16$. Note that a one-tailed test was used because the hypothesis is directional. The two-tailed test was also significant (p=0.017).

Of the 20 participants, 16 were rated as "good" or "moderate" by both examiners. Three were rated as "poor" by both examiners. Only one was rated "good" by one examiner and "poor" by the other. Cohen's Kappa was 0.0365, and the association between the ratings expressed in a 3x3 contingency table was significant by Fisher's exact test, p=.011.

The rating was significantly related to criteria for the diagnosis of visual stress proposed in a Delphi study by Evans et al [11] (see Table 4 of reference [11]). To pass the criteria participants needed two of the following three signs: 3 months' use of overlays, an increase in reading speed of 15% or more with an overlay, a Pattern Glare score of more than 3 with a mid spatial frequency grating. They also needed 3 of 6 typical symptoms. The signs are shown as present (1) or absent (0) in columns 4-6 of Table 1. The number of symptoms is listed in column 7. The three patients rated "poor" by both examiners did not meet the Delphi criteria for visual stress [11]. To analyse this contingency statistically, the examiners' rating was converted to a numerical value by summing a score of 2 for "Good", 1 for "Moderate" and 0 for "Poor" across both raters. The mean rating score for the individuals who passed the criteria for visual stress was 3.56 and for those who failed 1.0, two-tailed $t(18)=4.59$, p=.0002, d=1.73.

Overall, the mean difference in chromaticity between the two colorimetry assessments was 0.043 (SD 0.027), median 0.037. The difference was smaller for the 16 participants who were rated by both examiners as either "good" or "moderate": 0.037 (SD 0.022), median 0.033. For the three participants rated "poor" by both examiners (omitting the one participant who was rated "good" by one examiner and "poor" by the other) the difference in the chromaticity of the two colorimeter assessments averaged 0.083 (SD 0.011), median 0.086. The similarity between the two colorimeter assessments was significantly greater in the participants rated as "good" or "moderate" by both examiners than those rated as "poor" by both (two-tailed $t(17)=3.42$, p=.003, d=1.79). However, the number of "poor" participants was small. The chromaticities are shown in Figure 1.

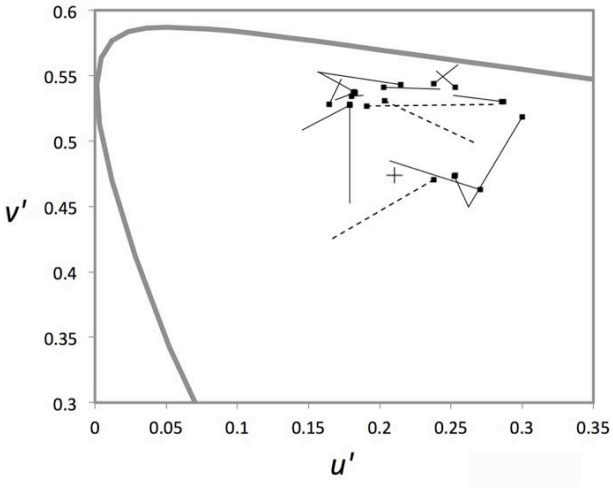

Figure 1. The chromaticities obtained in the first colorimeter assessment are marked by a point. They are connected by a line to the chromaticities from the second assessment. Broken lines represent the participants whose consistency was rated as poor by both examiners. In the first assessment saturation was constrained to 30, so the second assessment usually gave chromaticities that were more saturated (i.e. further from equal energy white, shown by the cross)

Given that the first colorimetry assessment was non-standard and restricted the saturation to 30, the two assessments are perhaps best compared in terms of the hue chosen rather than the combination of hue and saturation and associated chromaticity. The hue was identical in 7 of the 20 participants and differed by 30 degrees or less in 16 of the 20. The probability of the agreement occurring by chance was estimated numerically as follows. The pairing of one hue setting with another was randomised across the 20 participants, and the difference in hue angle squared (to remove the effects of the sign of the difference) and summed. The process was repeated 1000 times and only six of the 1000 repetitions gave a sum lower than that obtained from the correct pairing. A similar analysis of the chromaticity differences was also undertaken. The Euclidian distance in UCS between the two colorimetry assessments was obtained (by Pythagoras) when the pairing was randomised across participants 1000 times. None of the chromaticity differences obtained from such randomisation was as small as that obtained from the correct pairing.

4. Discussion

Under conditions in which the examiner was masked, the patients read randomly ordered common words more quickly with lenses tinted to match their colorimeter setting than with lenses that differed by 0.078 in chromaticity. Fourteen patients had not had prior experience with trial lenses. None had had experience with the particular lenses associated with their colorimeter setting. Reading randomly ordered words is artificial, and the numerical difference in reading speed is small, but the task has validity in predicting benefit from the use of coloured filters [5].

The examiners concurred in their assessment of the quality of the judgements given by their patients: 14/20 of the assessments were identical and 19 (all but one) were similar. The patients who passed the criteria for visual stress received higher ratings, supporting the criteria proposed by Evans et al [11]. The ratings predicted both the repeatability of the two assessments and the benefit from the tint. This suggests that examiners can infer the likely repeatability of patients' judgements during one colorimetry session. These judgements are likely to be of clinical use. It has been shown that the benefits of a coloured filter in increasing reading speed are greater in patients who give consistent results [16].

The repeatability of the two colorimeter examinations was similar to that obtained earlier by Suttle et al, [1] notwithstanding (1) the masked protocol; (2) the use of different models of colorimeter; (3) different assessment procedures, one restricting saturation; (4) a shorter but consistent interval between the two assessments. The similarity between the two studies is encouraging. For the "consistent" observers in the earlier study [1] i.e. those who chose the same or similar overlays, the chromaticity difference averaged 0.035. It averaged 0.053 for the "inconsistent" participants. The findings in the present study were similar: the chromaticity difference for participants rated as "good" or "moderate" by both assessors averaged 0.037, as compared with 0.067 for the remaining four participants.

The mean chromaticity difference for the consistent participants was averaged for the two studies (the present study and the previous study [1]) and was 0.036. The standard error of this mean was estimated from the standard deviation of data from both studies and was $0.026/\sqrt{40} = 0.0041$.

The cluster of points in a chromaticity diagram that are associated with improved clarity can be modelled most simply as a bivariate normal distribution in which the distribution of points is centred on an optimal chromaticity; i.e. their u' and v' coordinates are independently and normally distributed, see Figure 2.

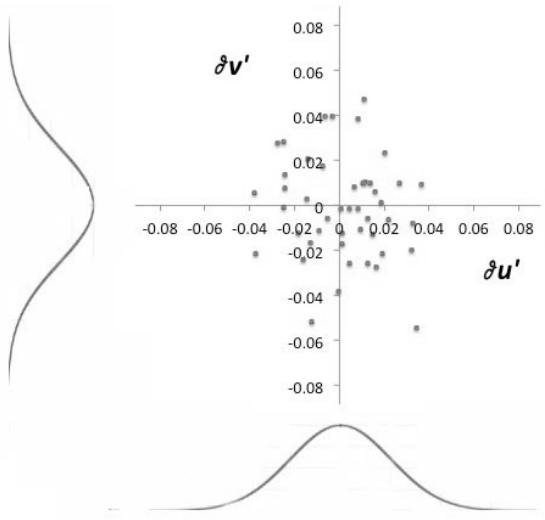

Figure 2. Distribution of hypothetical colorimeter adjustments centred on a single optimal chromaticity with a bivariate normal probability density. The optimal chromaticity is at the origin and the axes show the departure from optimum in units of u' and v'. The separation of any two points averages 0.036. The standard deviation of the points, σ, is therefore $0.036/\sqrt{\pi} = 0.020$

The average difference between two points in a bivariate normal distribution is the square root of pi times the standard deviation. The standard deviation of the u' and v' coordinates can therefore be estimated to be $0.036/\sqrt{\pi}$ with lower and upper confidence limits (estimated from the standard error of the mean given earlier) of 0.016 and 0.025 respectively. In the introduction reference was made to a chromaticity difference at which little benefit remains. The difference was 0.07 and this corresponds to more than three standard deviations.

The standard deviation of the u' and v' coordinates permits an estimate of the number of tints needed for a tinting system that will offer patients a sufficient range of tints. The chromaticity diagram can be most efficiently tessellated by hexagons, the centre of each hexagon representing the chromaticity of a trial lens. For a system to have a sufficient number of trial lenses, there should always be a lens available that is no more than (say) one standard deviation distant from the chosen chromaticity, as in Figure 3.

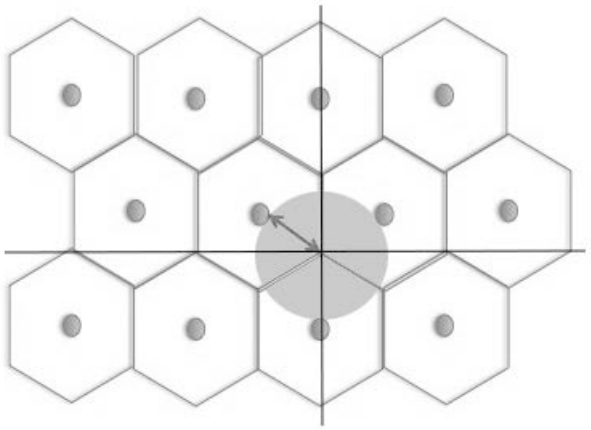

Figure 3. A UCS surface tessellated by hexagons. At the centre of each hexagon is the chromaticity of a trial lens. The scatter of chosen chromaticites is represented by the grey circle with its centre at the intersection of the horizontal and vertical axes, as in Figure 2. The chromaticities of the trial lenses closest in colour are one standard deviation from the centre of the scatter, as represented by the arrow

This will occur when the side of the hexagons is equal to one standard deviation and the area of each hexagon is therefore $3/2 \times \sqrt{3} \times 0.020^2 = 0.00104$. The area of the gamut available for use with conventional CR39 dyes that transmit more than 5% is given as 0.08 in reference [17]. Therefore $0.08/0.00104 = 77$ tints are required to cover this gamut with a resolution of 1σ. The estimate given above has confidence limits that may be estimated from the mean +/- 2 standard errors of the mean. The lower and upper confidence limits of the estimate of the number of tints required are thus 39 and 222 respectively. The estimates are similar to those from a previous study based on different methods and data [17].

Although a system with hexagonally tesselated tint chromaticites would be efficient, it might be difficult to maintain the necessary accuracy of each tint. It is therefore more realistic to use a larger number of trial lenses with greater tolerance. The chromaticities of the trial lenses in the Intuitive Colorimeter system show no gaps in the distribution of chromaticities that are larger than 0.020 [6,7].

The above analyses have used the CIE UCS diagram rather than a cone-opponent diagram. This use maintains continuity with previous reports of colorimetry. The approach is supported theoretically by the observation that both discomfort and its physiological correlate, the amplitude of the cortical haemodynamic response [9,18], are both affected strongly by the difference in CIE UCS chromaticity, and not so strongly with other measures of colour difference based upon cone contrasts, see Table 2 of reference [19].

Suttle et al [1] use their findings to suggest that either "the use of colour to alleviate discomfort or difficulty reading is not a valid approach, or that the use of colour is valid but the colour does not need to be precise." The first inference is not valid and the second depends on what is meant by precision. The findings presented here are entirely consistent with previous literature. They show that when colour improves reading speed it does so optimally only if within about 0.020 of the CIE 1976 UCS chromaticity chosen as providing "clarity and comfort" of vision.

Acknowledgements

Dr Kalvis Jansons provided useful comments.

Declaration of interest

The last author (AW) invented the Intuitive Colorimeter when he was employed by the UK Medical Research Council. He receives emoluments from the Council, and from Essex University based upon sales of the colorimeter and the Wilkins Rate of Reading Test, but not for the sale of coloured lenses. AW and PA have received honoraria for lectures on the use of coloured lenses in eye care practice. The other authors have no conflicts of interest to declare.

References

[1] Suttle CM, Barbur J, Conway ML. Coloured overlays and precision-tinted lenses: poor repeatability in a sample of adults diagnosed with visual stress. *Ophthalmic Physiol Opt*. 2017; 37: 542-548.

[2] Wilkins AJ, Evans BJW, Brown JA, et al. Double-masked placebo-controlled trial of precision spectral filters in children who use coloured overlays. *Ophthalmic Physiol Opt*. 1994; 14(4): 365-370.

[3] Wilkins AJ, Nimmo-Smith I, Jansons JE. Colorimeter for the intuitive manipulation of hue and saturation and its role in the study of perceptual distortion. *Ophthalmic Physiol Opt.* 1992; 12(3): 381-385.

[4] Wilkins AJ, Sihra N. A colorizer for use in determining an optimal ophthalmic tint. *Color Res Appl*. 2001; 26(3): 246-253.

[5] Wilkins A. *Reading Through Colour: How coloured filters can reduce reading difficulty, eye strain, and headaches.* 2003; Wiley:Southampton. 176pp.

[6] Wilkins A, Milroy R, Nimmo-Smith I, et al. Preliminary observations concerning treatment of visual discomfort and associated perceptual distortion. *Ophthalmic Physiol Opt.* 1992; 12(2): 257-263.

[7] Wilkins AJ, Sihra N, Myers A. Increasing reading speed by using colours: Issues concerning reliability and specificity, and their theoretical and practical implications. *Perception*. 2005; 34(1): 109-120.

[8] Wilkins AJ, Patel R, Adjamian P, Evans BJW. Tinted spectacles and visually sensitive migraine. *Cephalalgia*. 2002; 22(9): 711-719.

[9] Huang J, Zong X, Wilkins A, Jenkins B, Bozoki A, Cao Y. FMRI evidence that precision ophthalmic tints reduce cortical hyperactivation in migraine. *Cephalalgia*. 2011; 31(8).

[10] Coutts LV, Cooper CE, Elwell CE, Wilkins AJ. Time course of the haemodynamic response to visual stimulation in migraine, measured using near-infrared spectroscopy. *Cephalalgia*. 2012; 32(8).

[11] Evans BJW, Allen PM, Wilkins AJ. A Delphi study to develop practical diagnostic guidelines for visual stress (pattern-related visual stress). J Optom. 2016.

[12] Wilkins A. Overlays for classroom and optometric use. *Ophthalmic Physiol Opt.* 1994; 14(1): 97-99.

[13] Wilkins AJ. *A system for precision ophthalmic tinting: Manual for the Intuitive Colorimeter and trial lenses.* 1993; Cerium Visual Technologies: Kent.

[14] D'Ath PJ, Thomson WD, Wilkins AJ. Memory for the color of non-monochromatic lights. *Color Res Appl.* 2007; 32(1): 11-15.

[15] Wilkins, A.J., Allen, P.M.,Monger, L. and Gilchrist, J. Visual stress and dyslexia for the practising optometrist. *Optometry in Practice.* 2016; 17(2): 103-112.

[16] Wilkins A, Lewis E, Smith F, Rowland E, Tweedie W. Coloured overlays and their benefit for reading. *J Res Read.* 2001; 24(1): 41-64.

[17] Wilkins A, Sihra N, Nimmo-Smith I. How precise do precision tints have to be and how many are necessary? *Ophthalmic Physiol Opt.* 2005; 25(3): 269-276.

[18] Wilkins, A.J. A physiological basis for visual discomfort: application in lighting design. *Lighting Research and Technology*, 2016; 48: 44-54.

[19] Haigh, S.M., Barningham, L., Berntsen, M, et al. Discomfort and the cortical haemodynamic response to coloured gratings. *Vis. Res.*, 2013; 89: 47-53.