Color Vision Assessment - 3. An efficient, two-step, color assessment protocol

John L Barbur

Marisa Rodriguez-Carmona

Benjamin E W Evans

Affiliation:

Centre for Applied Vision Research, School of Health Sciences,

City, University of London, UK.

Corresponding author:

John L Barbur,

Professor of Optics and Visual Science,

City, University of London,

Email: J.L.Barbur@city.ac.uk

Phone: +44 2070405060

(Accepted for publication 10 November 2020)

ABSTRACT

Color vision tests and multi-test protocols in current use often fail to detect small changes in red/green (RG) and yellow/blue (YB) color vision due to poor sensitivity. The tests also have low specificity. In this study we examine how improved understanding of within- and inter-subject variability in RG and YB color vision and accurate assessment of the differences in color thresholds between the least-sensitive, age-matched normal trichromats and the least-affected deutans and protans can be used to design an efficient Color Vision Screener (CVS) test. To achieve this objective, we examined two extensive data sets from earlier studies and carried out new experiments to provide better estimates of within-subject variability in color thresholds and to validate the CVS test. The data sets provide essential information on inter-subject variability, the effects of normal aging on RG and YB thresholds[1] and the spread in RG color thresholds in deutan and protan subjects[2].

A statistical model was developed to optimise the parameters of the CVS test and to predict the limits of what can be achieved in color assessment. The efficiency and repeatability of the CVS test was then assessed in 84 subjects. The results match model predictions and reveal close to 100% test efficiency. The test takes between 140s to 160s to complete and has close to 100% repeatability. An efficient, 'two-step' protocol based on the initial use of the CVS test followed by full color assessment in only those who fail the CVS test is also described.

Keywords: color assessment, red-green color deficiency, yellow-blue color deficiency, acquired color deficiency, CAD test, color vision screener

1. INTRODUCTION

47 Color discrimination is arguably the most sensitive attribute of vision with only gradual changes
48 as a result of normal aging[1, 3-6]. Congenital color deficiency affects mostly RG color vision in ~
49 8.8 % of males and ~ 0.4 % of females[7, 8]. The severity of RG loss in congenital color deficiency
50 varies from almost normal chromatic discrimination sensitivity to complete absence of RG color
51 vision[9]. Acquired color deficiency arises mostly as a result of diseases of the retina and the optic
52 nerve with preferential loss of YB chromatic sensitivity during the earliest stages of disease[10, 11].
53 Preferential loss of RG chromatic sensitivity is less common, but has been reported in patients
54 with Duchenne muscular dystrophy[12] and congenital aniridia[13]. Although acquired deficiency is
55 most common in older people, younger subjects can also be affected, particularly when systemic
56 diseases are involved[14, 15].
57 Color vision carries useful information that can greatly enhance visual performance[16, 17]. In
58 some visually demanding occupations, the use of color signals is safety-critical and reliable
59 assessment of severity of color vision loss becomes important[18-20].
60 Screening for normal, trichromatic color vision is most commonly carried out using one of the
61 several editions of the Ishihara (IH) plates and other pseudoisochromatic tests followed by
62 minimum hue discrimination tests such as the Farnsworth D15, City University test, lantern tests
63 and RG color matching, usually based on Rayleigh's yellow match[21, 22]. Although several
64 parameters affect the outcome of each of the above color assessment tests, the wavelength
65 separation, $\delta\lambda_{max}$, between the peak spectral responsivities of L and M cones contributes most to
66 the level of RG chromatic sensitivity that can be achieved[23]. In normal trichromats, $\delta\lambda_{max}$, is ~ 28
67 nm. The least-affected deutans can approach a maximum separation of ~ 10 to 12nm, but the
68 least affected protans can only achieve a maximum of 6 nm[24, 25]. The relationship between the
69 severity of RG color vision loss and $\delta\lambda_{max}$ is also very nonlinear with only modest improvements

in sensitivity when $\delta\lambda_{max}$ exceeds 12 nm[23]. Some subjects diagnosed genetically as RG dichromats continue to exhibit residual RG chromatic discrimination sensitivity which has been attributed to other factors such as differences in the optical densities of L and M cone pigments[26]. In the case of anomalous trichromats, other parameters such as the large variation in the relative numbers of L and M cones in the retina or the presence of variant cone pigments in addition to normal pigments in female carriers of color deficiency can also affect the outcome of color assessment tests[27]. The correlation between detection of and predictions of expression of variant pigment genes based on genetic tests and the deficiency of color discrimination remains poor[28]. Whilst the anomaloscope and CAD assessments of the type of color deficiency are in very close agreement, the correlation between the parameters of the yellow match and the subject's red / green chromatic sensitivity is poor[29]. On those rare occasions when subjects classed as normal on the anomaloscope, within the variability associated with the parameters of the yellow match, are diagnosed with congenital deficiency on the CAD test, other parameters may be involved. Such outcomes are possible when in addition to small shifts in the peak spectral responsivity of L or M cone pigments, changes in their optical densities are also involved. The latter can cancel out the shift in the midpoint of the yellow match caused by the presence of variant pigments, making it difficult to diagnose the presence of mild congenital deficiency[30]. In such cases, genetic testing can provide useful, definitive information on the presence or absence of congenital RG deficiency[9], but less so on the expected severity of color vision loss.

In this series of three papers we examine the outcome of the most common color assessment tests employed in the clinic to detect early changes in chromatic sensitivity, and also in occupations to screen for congenital and / or acquired loss of color vision. In the first paper[31], we examine the stimulus related visual signals a subject can make use of to pass color arrangement tests such as

93 Farnsworth D15. We show conclusively that subjects can make use of RG, YB and even
94 luminance contrast signals to judge minimum perceived 'color' differences between adjacent
95 caps and to carry out the arrangement test with no errors. We also show that dichromats and even
96 rod monochromats can carry out the cap arrangement tasks with no errors when the caps are
97 presented in two subgroups that preserve monotonic changes in either S-cone or rod signals. In
98 the second paper[22], we examine the statistical outcomes of the most common clinical and
99 occupational color assessment tests and calculate the efficiency of frequently used, single- and
100 multi-test protocols with emphasis on sensitivity and specificity. Many multi-test protocols
101 employ three or even four tests to reach an outcome. Analysis of color assessment outcomes[21, 22,
102 32, 33] and comparison of different tests and protocols show that none of the commonly used, color
103 assessment protocols approach high sensitivity and specificity. These tests also fail to quantify
104 reliably the severity of color vision loss. The overall efficiency of both single and multi-test
105 protocols remains low since the tests can only achieve high sensitivity at the expense of
106 specificity[34]. The results reported in the second manuscript also reveal the often unjust outcomes
107 of multi-test protocols in current use. When the assessment protocols are designed to pass all
108 normal trichromats, many subjects with congenital and / or acquired color deficiency also pass[22,
109 31], some with severe loss of RG color vision and at the same time, many less severe subjects
110 fail[22].
111 Complete color assessment requires full isolation of color signals and also selective stimulation
112 of RG and YB chromatic mechanisms. High sensitivity is required to separate the least-sensitive
113 normal trichromats from the least-affected deutans and also to detect the smallest changes in
114 color vision during disease progression or treatment. The ideal test should be able to detect
115 acquired loss in patients with congenital RG deficiency, classify accurately the applicant's class

of color vision with 100% sensitivity and specificity and also quantify reliably the severity of RG and YB loss.

No current conventional test or protocol fulfills all of these requirements, mainly because the tests fail to isolate the use of only RG or only YB color signals, do not eliminate the potential use of other cues and fail to achieve close to 100% sensitivity and specificity[22, 31, 35]. When used appropriately, some color threshold tests developed more recently, such as CAD[23] and the Cambridge Color Test (CCT)[36, 37] come close to achieving the ideal test requirements. These tests are, however, disadvantaged by the high costs of using expensive equipment which requires calibration. As a result, they are not commonly available and the thorough examinations needed to classify accurately the class of color vision and the severity of loss take long to complete (usually between 12 and 15 minutes on the CAD test).

The third manuscript in the series describes an efficient 'two-step' color assessment protocol designed to overcome many of the disadvantages associated with conventional color assessment protocols[22].

Step 1 requires screening for normal RG and YB loss of chromatic sensitivity. The ideal color vision screener must be inexpensive, take little time to administer and must pass all normal trichromats and fail those with congenital and / or acquired color deficiency.

Step 2 requires all applicants who fail the screener to take the full CAD test (or equivalent). If rapid 'screening' for congenital and / or acquired color deficiency can be carried out with 100% sensitivity and specificity, only ~ 6% of all those tested require full color vision assessment. This estimate assumes equal numbers of males and females being tested and ~ 2% prevalence of acquired color deficiency. The latter may be even higher than 2%, particularly in older subjects[38]. A 'two-step' protocol based on the initial use of a color vision screener test followed

by full color assessment in only those applicants who fail the screener test has many advantages, but this requires the availability of an inexpensive and rapid screener with close to 100% test efficiency.

In this manuscript we describe a new CVS test which comes close to meeting the desired requirements. The stages involved in this test have been modelled statistically with parameters derived from measured data in normal trichromats and in subjects with congenital color deficiency. The model allowed optimum selection of remaining test parameters. When the least-sensitive normal trichromats are compared with the least-affected deutan subjects, the model predicts close to 100% sensitivity and specificity. The less sensitive, least-affected protan subjects always fail the CVS test. We report preliminary color assessment results, including the outcome of the new CVS test in 84 subjects examined before the imposition of restrictions on face-to-face examinations as a result of the pandemic. The results are completely consistent with model predictions. In addition to predicting the high efficiency of the CVS test, the statistical analysis carried out also reveals the limits of what one can achieve experimentally as a result of inter-subject and within-subject variability.

2. METHODS

2.1 Participants

The study participants attended the Advanced Vision and Optometric Tests (AVOT) clinic at City, University of London. Color assessment was carried out in each subject using IH, CAD, and the Nagel anomaloscope. Although the participants attended the clinic for color vision assessment, informed consent was obtained from each participant to allow anonymous inclusion of their results in our research investigations. The analysis carried out in this study includes two

sets of data measured over several years in addition to data from experiments designed specifically for this study. The first set of data examined the effects of normal aging on RG and YB color vision in normal trichromats. The data provided upper-normal, threshold limits as a function of age for both RG and YB color vision and a measure of inter-subject variability. The study involved 215 females and 197 males, age range 4 to 90 years[1]. The second study measured the loss of RG sensitivity in color deficient subjects in relation to age. In total, 268 deutans (age range: 9 to 58 years), and 132 protans (age range: 7 to 58 years) were examined. The rankings of RG thresholds in the deutan and protans groups have been reported previously[38]. Within-subject variabilities in RG and YB color thresholds were estimated in three normal trichromats, labelled as subjects S1, S2 and S3 in Figure 4(a). The analyses of age-matched thresholds in normal trichromats and in least affected deutans and protans and the estimated within- and inter-subject variabilities aided the development of the new screener test and the formulation of the statistical model needed to understand and to predict the sensitivity and specificity of the test. Finally, the efficiency and repeatability of the screener were assessed experimentally in a separate, multi-center study involving a number of CAD test centers who were able to include the CVS test in their own color assessment procedures based on CAD. The same informed consent was employed. 84 subjects were investigated before the imposition of restrictions on face-to-face examinations as a result of the pandemic, but it is our intension to resume the study as soon as it is safe to do so. The sample included 27 normal trichromats (age range: 17 to 51), 41 deutans (age range: 18 to 56), 14 protans (age range: 19 to 52) and two deutan subjects who also had acquired loss of color vision (age 27 and 32 years). The study followed the tenets of the Declaration of Helsinki and was approved by the Research and Ethics Committee of City, University of London.

185

186

187 **2.2 The Color Vision Screener test**

188 The CVS test was designed to be rapid, simple and easy to use. This test has not been described
189 previously. The test employs a 'two-alternative' forced choice procedure which links two
190 keyboard buttons (numbers 7 and 9, Fig. 2 (b)) with the 'top-left' and 'top-right' motion
191 directions of a color-defined stimulus buried in dynamic luminance contrast noise (Fig. 2 (a)).
192 The test includes an interactive learning step followed by a short test with suprathreshold trials
193 which all subjects must pass before proceeding with the full test. The stimulus parameters are
194 similar to those developed for the CAD test[38, 39] to ensure that the subject can only make use of
195 either RG or YB color signals to produce a correct response, but the testing procedure is much
196 simplified. There are 88 trials with 'signal colors' to assess the normal functioning of RG and
197 YB chromatic mechanisms and 22 suprathreshold trials when RG and YB colors are added to
198 stimuli defined by luminance contrast to ensure that all subjects, including dichromats are able to
199 detect the moving test stimuli and to produce a correct response (Fig. 2, (d)). The suprathreshold
200 trials help reinforce correct responses, particularly in subjects with color deficiencies and provide
201 a means of assessing the subject's response reliability. The latter is determined by the number of
202 errors the subject makes (i.e., the error score, E) in suprathreshold trials when the probability of a
203 correct response is expected to be 1. The subject's response reliability is classified as follows:
204 'Excellent': E=0; 'Good': E=1; 'Poor': $3 \geq E \geq 2$; 'Unusable': $E \geq 4$
205 'Unusable' is often linked to a total lack of attention during the test. Before proceeding with the
206 color vision screener test, each subject has to pass a 'short test' with suprathreshold stimuli. A
207 'Pass' on this test ensures full understanding of test requirements, the ability to see motion and to

208 respond appropriately to moving stimuli. A 'Pass' also rules out inability to learn the task or the
209 presence of extremely poor vision as a result of severe retinal or systemic disease affecting
210 stimulus attributes such as motion which could also be used to explain an 'Unusable' outcome.
211 The diagnosis of normal RG, normal YB or the presence of color deficiency is based on the
212 correct response scores to the 44 RG and 44 YB signal colors (Table 2). The scores are governed
213 by a binomial distribution and determined entirely by the mean probabilities of a correct
214 response to RG (p_{rg}) and YB (p_{yb}) stimuli. These probabilities are in turn determined by the
215 subject's sensitivity to color signals, normally quantified by measuring the subject's RG and YB
216 color thresholds.
217
218 2.**3 The CAD test**
219 The CAD test is based on results of studies on camouflage which revealed how the use of color
220 signals can be isolated using dynamic luminance contrast noise[39]. The colored stimuli employed
221 have a space-averaged luminance that matches that of the flickering field. This is only true for
222 the standard CIE observer[40], but the use of dynamic luminance contrast noise ensures that even
223 dichromats cannot make use of residual luminance contrast signals to detect the direction of
224 motion of the colored stimulus [23, 41]. The standard CAD test uses 16 interleaved color directions
225 specified in the CIE 1931 (x, y) color space. Following each presentation, the subject's task is to
226 press one of four buttons, to indicate the direction of motion of the color-defined
227 stimulus. The CAD test has evolved over several decades and is currently in use in many
228 visually demanding occupations. One important advantage of the CAD test is the introduction of
229 CAD units. The standard normal CAD units for RG and YB color vision are based on mean
230 color thresholds measured in 330 healthy, young, normal trichromats[34]. For convenience all color

thresholds measured in this study are expressed in CAD units. One CAD unit defines the standard CAD observer. In comparison, a RG or YB threshold of 4 CAD units indicates that the applicant requires four times as large color signal strength, when compared to young normal trichromats, to achieve the same level of chromatic discrimination sensitivity. The measured thresholds are directly proportional to the cone contrast generated by the colored stimulus[23].

3. RESULTS AND DISCUSSION

3.1 The Ishihara test

The 'two-step' protocol requires the initial use of a sensitive color vision screener test to separate subjects with congenital deficiency from normal trichromats. Before the development of the CVS test, we searched for the ideal color vision screener, and with this aim in mind, we examined the sensitivity of many of the conventional color vision tests and found that the IH 38-plates edition had the highest sensitivity, when no errors are allowed on plates 1 to 25 (Figure 1(a)). The high sensitivity is, however, achieved by sacrificing specificity with just over 18% of normal trichromats also failing the test. On the assumption that equal numbers of males and females are tested for color deficiency, just under 22% of all applicants fail and require full color assessment when using IH as a screener. This percentage is 3.6 times larger than what can be achieved with a screener test of close to 100% sensitivity and specificity. There are also other disadvantages of using the IH test either as a screener or for full color assessment. The IH test does not screen for YB color vision and cannot therefore separate acquired from congenital loss or acquired loss in patients with congenital RG deficiency. Since acquired loss of color vision affects predominantly YB chromatic mechanisms, subjects with acquired loss may pass undetected. The test misclassifies some deutans and protans with many left indeterminate (Fig. 1(a)). It is also well established that determined applicants can learn to use other cues to pass the

IH test, even when severely deficient[42]. Finally, the correlation between the number of plates the subject fails and the severity of RG color vision loss is very poor[34]. Although in general subjects with reduced chromatic sensitivity tend to make more errors on the IH test, the correlation between the number of errors the subject makes and the severity of RG loss remains poor (Fig. 1 (b)). When four or fewer errors are accepted as a pass in order to pass all normal trichromats, many subjects with RG color deficiency also pass, some with severe loss of RG color vision (see highlighted area in Fig. 1 (b)).

3.2 Age dependence and inter-subject variability

Fig. 3 (squares) shows how RG color thresholds vary as a function of age in healthy subjects with normal trichromatic color vision[1]. The green and red circles show data measured in subjects diagnosed with congenial deutan or protan color deficiency, using the CAD test. Those with congenital deficiency vary in severity of loss from close to normal upper thresholds, to complete absence of color vision with thresholds limited only by the gamut of the visual display (usually around 34 CAD units[23] for the stimulus conditions employed in the test). For completion, the same subjects were also examined for RG congenital color deficiency and classified as deutan- or protan-like using the Nagel anomaloscope. The CAD and Nagel classifications in this study were found to be in 100% agreement.

The results shown in Fig. 3 capture the effects of normal aging as well as the inter (σ_I) and within-subject (σ_w) variabilities for RG color vision. Very similar results have also been found in normal trichromats for YB color vision, but with a slightly more rapid increase in thresholds with increasing age[1]. The CAD test employs interleaved two-up / one-down staircases which yield the RG and YB color signal strengths that correspond to 0.71 probability of a correct response[23].

277 Although the mean probability is 0.71, the within subject variability causes a spread in this
278 probability and consequently a spread in repeated thresholds with standard deviations (σ_w) in the
279 range ~ 5 to 15 percent of the mean threshold. The CVS test employs a fixed color stimulus
280 strength set just below the upper normal limit for the applicant's age (Figure 3). The outcome of
281 the screener test is determined entirely by the applicant's probability of a correct response to the
282 stimulus strength employed selected for the test.
283
284 **3.3 Within- and inter-subject variability**
285 Typical results derived from repeated RG and YB measurements are shown in Fig. 4 (a). Fig. 4
286 (b) shows inter-subject variability which also includes the spread caused by within-subject
287 variability. The measured thresholds are approximately normally distributed and not
288 unexpectedly, the measured inter-subject variability is significantly larger, with σ_I values in the
289 range 21 to 25 % of the mean (Fig. 4 (b)). These data are important in predicting the expected
290 outcome of the CVS test in normal trichromats and also in congenital and acquired color
291 deficiency. A number of observations can be made based on the data shown in Fig. 3 and 4.
292 a. The 'least-sensitive' normal trichromats with thresholds at the upper normal limit (UNL)
293 of $2.5\sigma_I$ overlap with the 'least-affected' deutans. Subjects with thresholds close to this
294 limit represent only a small percentage within both the normal and the deutan classes.
295 Only subjects at the extreme ends of their inter- and within-subject variability (Fig. 4 (a,
296 b)) approach the UNL, indicated by the upper, red, dotted line in Fig. 3.
297 b. The 'least-affected' protan-like subjects have thresholds much higher than the UNL (Fig.
298 3, red circles) and are therefore guaranteed to fail the CVS test for any RG color signal
299 strength below the UNL.

c. The within-subject variability is subject-specific with values of ~ 10% of the mean threshold (Fig. 4 (a)) and virtually eliminates the threshold difference between the 'least-sensitive' normal trichromats and the 'least-affected' deutans.

These observations suggest that a color signal strength just below the UNL for the 'least-sensitive' normal trichromats, would be appropriate for use in the CVS test to separate the latter from the least-affected deutans. A color signal strength equal to $\mu_I + 2\sigma_I$, where μ_I represents the median threshold for the subject's age (indicated by the black dotted line in Fig. 3) and, σ_I, the inter-subject variability, was selected for use in the CVS test.

3.4 CVS test results - preliminary findings

Table 1 summarises preliminary CVS test classification outcomes measured in 84 subjects (section A). Section B shows the results of a short repeatability study based on 24 repeated CVS tests in six normal trichromats and three deutans. The results show that all normal trichromats pass both the RG and the YB assessments. All protan and the deutan subjects fail the RG assessment, but they all pass the YB. The repeatability study shows 100% outcome repeatability for deutan subjects for both RG and YB assessment and 100% RG and 99.3% YB repeatability in normal trichromats. Only one normal trichromat fails the YB assessment in one, out of 24 trials.

3.5 Statistical model to predict the outcome of the CVS test

The screener test employs a two-alternative, forced-choice procedure with constant, RG and YB color signal strengths. The correct response scores are governed by a binomial distribution and determined by the mean probability, p, of a correct response. The latter is in turn determined by the subject's sensitivity to RG or YB color signals. Fig. 5 (a) shows the expected distribution of

RG CAD thresholds in a 'least-sensitive' 40-year-old normal trichromat (black) and the least-affected deutan (green), when the within-subject variability is assumed to be 10% of the mean threshold. The mean values of these distributions indicate 71% probability of a correct response. The probability of a correct response, p, was also measured as a function of color signal strength in normal trichromats and in subjects with congenital RG deficiency. Figure 5 (c) shows the probability of a correct response as a function of color signal strength in a least-sensitive, normal trichromat (black diamonds) and the model prediction for a least-affected deutan (green squares). The Weibull functions fitted to the data were of the form, $P(s) = 1-\exp(-(s/a)^b)$. *s* represents the color signal strength and *a* and *b* are the constants needed to provide the optimum fit. The curves were fitted using the Gauss-Newton analytic method in JMP (SAS Institute Inc.) program. The deutan data are based on the assumption that the within-subject variability is 10% of the mean threshold. The fitted functions enable us to predict the mean color signal strengths for 71% correct response which corresponds to the thresholds measured using the two-down, one-up staircase procedure[43] in the CAD test. In order to predict the outcome of the CVS test, we need to calculate the probability of a correct response, p, for the least-sensitive normal trichromat and also for the least-affected deutan and protan subjects, for the color signal strength of, $\mu_I + 2\sigma_I$, employed in the CVS test.

The following is a statistical analysis of expected outcomes for the CVS test. Fig. 3 also provides age data for the UNL to describe the least sensitive normal trichromat. The experimentally established normal, age-matched limit reflects the measured inter-subject variability[1] and was originally computed for $\mu_I + 2.5\sigma_I$.

3.6 Example calculations for a 40-year-old subject

For illustration purposes, we have chosen data for a subject, 40 years of age, but the same analysis is valid for any age. The only assumption made is that the within subject variability (σ_w) is 10% of the subject's mean color threshold, $\sigma_w = 0.1T_m$, where T_m represents the subject's mean threshold. This assumption is based on the measured within-subject variabilities shown in Fig. 4 (a). The key to understanding why the CVS test approaches 100% efficiency is the realisation that the mean threshold for the least sensitive normal trichromat is not at the UNL because of within-subject variability (Fig. 4 (a)). Since there are very few subjects ending up with a threshold $2.5\sigma_I$ above the mean[1], we have decided to extend the analysis to all subjects with thresholds $2\sigma_I$ above the mean (μ_I). It follows that $\mu_I + 2\sigma_I$ must therefore equal, $T_m + 2\sigma_w$, where T_m represents the subject's mean threshold for 71% probability of a correct response and σ_w is the standard deviation of subject's repeated measurements. Since $\sigma_w = 0.1T_m$, $T_m = (\mu_I + 2\sigma_I)/1.2$. This equation provides us with the mean color signal strength for 71% probability of correct response on those few cases when the subject's measured threshold is close to $\mu_I + 2\sigma_I$. Similarly, the deutan subjects with measured thresholds close to the UNL in Fig. 3 (green symbols) must have a mean threshold, T_m, equal to $\mu_I + 2.5\sigma_I + 2\sigma_w$, where σ_w represents the standard deviation which describes repeated thresholds in the deutan subject. Since σ_w is taken to be 10% of the subject's mean threshold, it follows that $T_m = (\mu_I + 2.5\sigma_I)/0.8$.

This analysis enables us to predict the mean color signal strengths (for 71% correct response score) for the least-sensitive normal trichromats and the least affected deutans, based on their expected mean color signal thresholds. The model assumes the same probability values, but the corresponding color signal strengths are multiplied by the constant needed to ensure that the 71% probability of a correct response corresponds to the mean thresholds predicted for the least-sensitive normal and the least-affected deutan. The optimum Weibull function parameters are

again evaluated and the results plotted for the normal trichromat (black diamonds) and for the deutan subject (green squares) in Fig. 5 (c). The horizontal dotted line if Fig. 5 (c) indicates the 0.71 probability of a correct response and the vertical dotted red line shows the color signal strength employed in the CVS test for a 40 years old subject (i.e., $\mu_I+2\sigma_I$). The probability of a correct response for the color signal strength employed in the CVS test can now be calculated for both the least-sensitive normal trichromat as well as for the least-affected deutan. The computed values are then corrected for the chance probability of a correct response and the corresponding binomial probability distribution functions with the predicted p-values for normal and deutan subjects calculated.

3.7 Expected outcomes for the least-sensitive normal trichromats and for the least-affected deutans

The binomial functions showing the probability of, n, correct responses (PDF) and the probability of n or fewer correct responses (CDF) are shown in Fig. 5 (b) and (d) for the least-sensitive normal and for a least-affected deutan subject, respectively. The plots indicate that based on this analysis, the CVS test can achieve close to 100% sensitivity and specificity. The computations assume that the within subject variability, σ_w, is 10% of the mean threshold (Fig. 5 (c)). Since the within-subject variability, σ_w, affects the expected probability of a correct response for the color signal strength employed in the CVS test, which in turn affects the corresponding binomial predictions, it is of interest to establish how the binomial predictions change, if the computation is based on measured probability of correct response data, with no assumptions made about within-subject variability. The corresponding percentage correct scores

390 and the measured data points for a least-affected deutan are shown in Fig. 5 (e), together with the
391 corresponding binomial predictions shown in Fig. 5 (f). The same analysis was carried out based
392 on measured data for a least-affected protan subject (Fig. 5 (g) and h)). Not unexpectedly, the
393 'least-affected' protans have thresholds much higher than the color signal strength employed in
394 the CVS test and as a result the test efficiency is close to 100%.

395 Table 2 summarises the predicted outcomes of the CVS test based on the measured data for
396 normal trichromats and for the least affected deutan and protan subjects. The analysis was
397 carried out separately based on both model assumptions and also on data measured for least-
398 affected deutan and protan subjects. The results show shows that all normal trichromats and all
399 protan subjects are classified correctly and that 97.5% of the least affected deutans (~ 5% of the
400 total deutan population) are also classified correctly. This means that only 0.13% of deutans are
401 expected to pass as normal trichromats.
402 In practical terms, less than 5% of subjects within each category can be described as 'least-
403 sensitive' normal trichromats and 'least-affected' deutans with thresholds close to the UNL. This
404 means that the number of deutans that can potentially be misclassified is extremely small. These
405 findings suggest that the CVS test has close to 100% test efficiency.
406
407 The high efficiency of the screener makes possible the introduction of a 'two-step' protocol for
408 use both in the clinic and in occupations. The new protocol enhances greatly the efficiency of
409 color assessment by reducing the number of subjects who require full color assessment using
410 CAD or equivalent tests to only 6% of all applicants. This estimate assumes equal numbers of
411 male and female applicants and 2% prevalence of acquired loss of color vision. The CVS test is
412 also available as a Menu option in the CAD test.

413 The screener has two built in limitations, it does not classify the patient's class of color vision
414 deficiency and does not quantify the severity of loss. These limitations confer some advantages
415 in that only small chromatic displacements are involved and accurate calibration of the computer
416 display is not a strict requirement. Small changes in background luminance and chromaticity do
417 not affect significantly the measured RG and YB color thresholds[23, 44]. The aim is to make the
418 CVS available for use on home computers with only minor adjustments to the computer display,
419 such as the selection of the sRGB color mode and the need for low mesopic ambient lighting
420 during the test. The use of the test on home computers will greatly expand color vision
421 assessment within schools and visually-demanding occupations. The test may prove extremely
422 useful clinically to detect the presence of significant changes in YB and RG color vision as a
423 result of early-stage diseases of the retina and the optic nerve with immediate advantages to
424 neuro-ophthalmologists in online consultations. In addition, the color vision screener can also be
425 used regularly by patients on home computers to monitor the progression of vision changes
426 caused by diagnosed systemic diseases such as diabetes or to detect the earliest significant
427 changes in color vision that, in some patients, can precede the clinical diagnosis of disease.
428
429 **4 CONCLUSIONS**
430 The new CVS test is rapid and easy to use by both children and adults. The CVS test detects both
431 congenital and acquired loss of chromatic sensitivity, provided the subject's thresholds fall
432 outside the upper, age-specific limits that describe normal RG and YB color vision. A knowledge
433 of within- and inter-subject variability and the use of age-specific color signal strengths make it
434 possible to separate the least-sensitive normals and the least affected deutans and protans with
435 close to 100% test efficiency and repeatability. This makes the new screener ideal for use in the

proposed ‘two-step’ color assessment protocol by providing rapid and accurate color screening to reduce the number of applicants needing full assessment to only ~ 6% of all the subjects tested.

Acknowledgements

This work was supported by the Colt Foundation and City, University of London. We also acknowledge the UK Civil Aviation Authority and The Rank Prize COVID-19 student support fund.

We are particularly grateful to Jack Werner for his valuable criticism and suggestions, and to Chris Hull, Vilhelm Koefoed and Tim Carter for their constructive comments and stimulating discussions.

The international CAD centers who are participating in the multi-center CVS validation study:

- Centreline Aviation – London, UK - Dr Chris King
- Deakin University - Melbourne, Australia - Dr Amanda Douglass
- Dr Vio & Partners - Hong Kong - Dr Steven Ho
- Livingstone Clinics - Melbourne, Australia - Dr Elizabeth Livingstone
- US Navy Refractive Surgery Center – San Diego CA, USA - Dr Vilhelm Koefoed
- Kharkevich Institute, Russian Academy of Sciences - Moscow, Russia - Professor Galina Rozhkova
- Sanjeevan Clinic – Mumbai, India - Dr Rohan Goyal
- Medizinisches Zentrum - Stuttgart Airport, Germany - Dr Sabine Roelcke

ORCID

John L Barbur
http://orcid.org/0000-0002-2187-5004

Disclosures:

The CVS test was first described[45] at ARVO 2019.

Marisa Rodriguez-Carmona,	City, University of London, None.
Benjamin Evans,	City, University of London, None.
John Barbur,	Employee of City, University of London. Company Director for COL Ltd. (A spin-out company set up by City, University of London to develop and market AVOT tests)

Figures and captions

(a)

Classification outcome based on zero errors (using the first 25 plates of the Ishihara 38 plates edition)				
% classified as:	N (336)	D (705)	P (319)	T(3)
N	81.85	0.71	0.63	100
CVD	18.15	99.29	99.37	0.0
D	0.00	60.14	8.20	0.0
P	0.00	0.14	38.49	0.0
Indet	18.15	39.71	53.31	0.0
T	0.0	0.0	0.0	0.0

SH 38 plates version (2-25 plates)

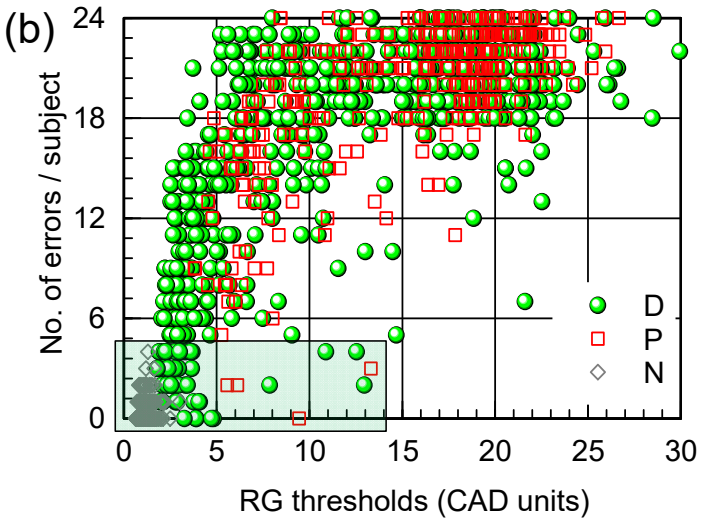


Fig. 1. Classification outcomes based on zero errors using the first 25 plates of the Ishihara 38 plates test edition (a). The results are based on 1363 subjects examined with Ishihara, Nagel anomaloscope and CAD. In the absence of acquired loss of color vision, there is close to 100% agreement between CAD and the anomaloscope in classifying the subjects as normal or RG deficient. This Ishihara test protocol yields the highest sensitivity in detecting subjects with congenital RG color deficiency. Almost all deutans and protans fail the test, but just over 18% of normal trichromats also fail. Only 60% of deutans and 38.5% of protans are classified correctly, the remaining subjects are either misclassified or indeterminate (i.e., classification not possible). Section (b) shows the relationship between the severity of RG color vision loss and the number of errors a subject makes on plates 2 to 25 of the Ishihara test. Normal trichromats are plotted as grey diamonds, protans as red squares and deutans as green circles. Exceptionally, normal trichromats can make up to a maximum of four errors. The shaded area along the axis shows the subjects who pass with four or fewer errors (modified from Barbur and Rodriguez-Carmona 2017)[23].

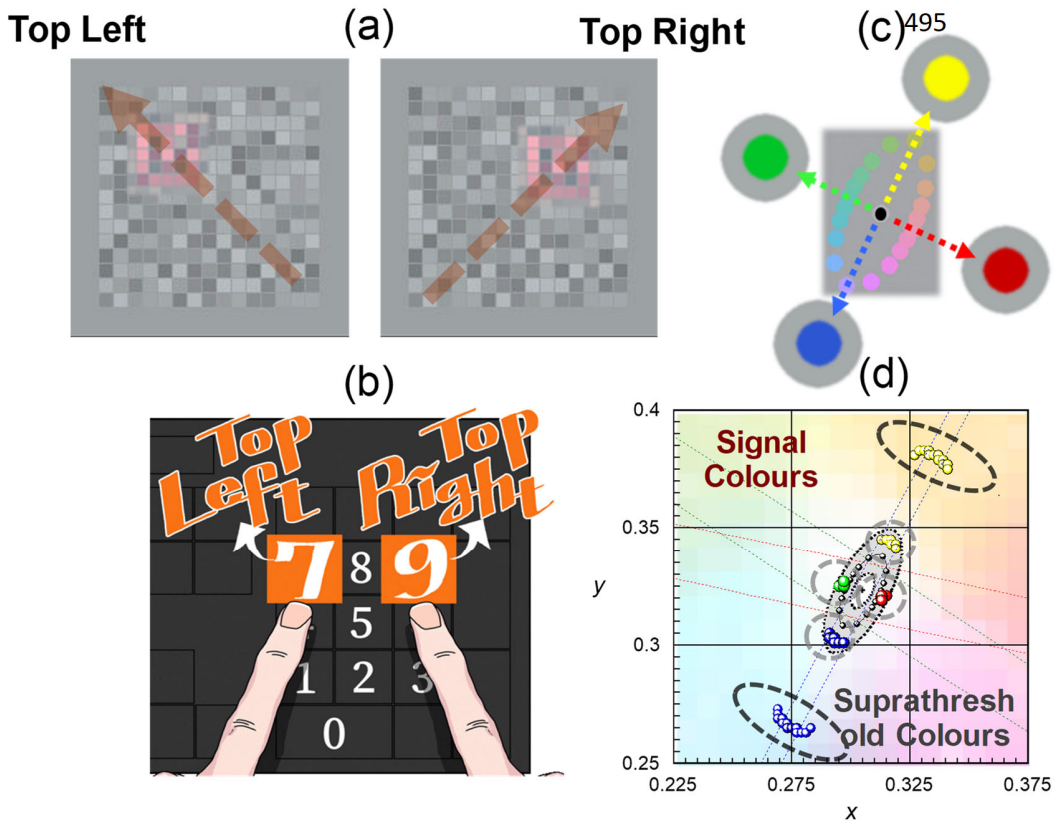


Fig. 2. Illustration of the motion directions (a) associated with the colored stimuli (c), the subject's task (b) and the colors employed in the CVS test (c, d). During each trial the chromaticity of the threshold 'signal colors' rotates smoothly over a range of angular directions that isolate the RG and YB chromatic mechanisms. The 'suprathreshold colors' consist of combined suprathreshold colors and luminance contrast signals to ensure detection by all subjects, including dichromats. In total, the test employs 110 stimulus presentations (i.e., 88 signal colors and 22 suprathreshold colors).

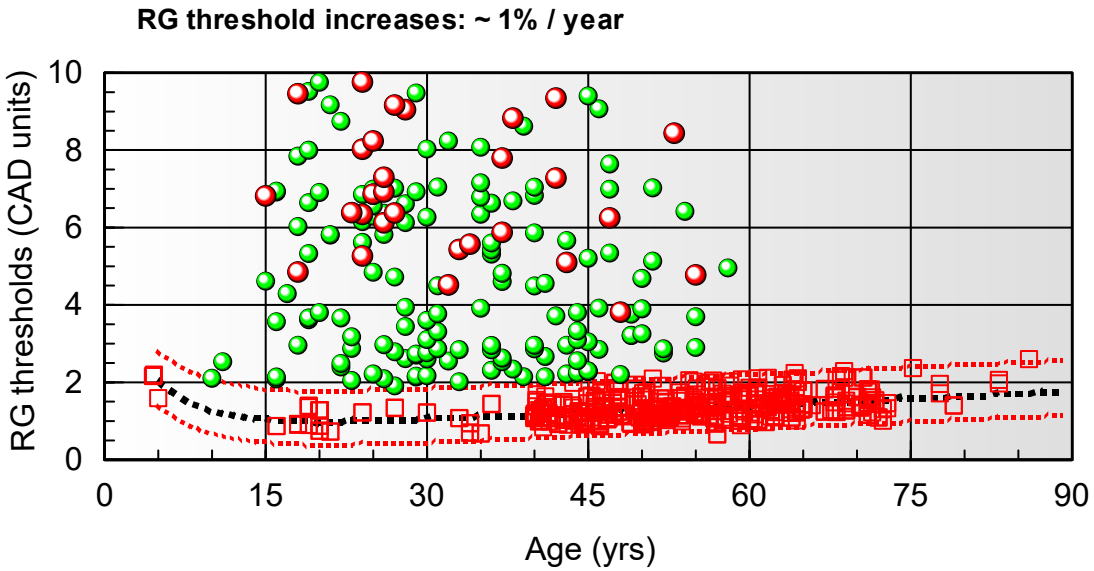


Fig. 3. RG color thresholds in healthy, normal trichromats[1] and in subjects with congenital RG color deficiency, plotted as a function of age. The dotted black curve indicates the change in mean threshold with age whilst the upper dotted, red curve indicates the UNL (i.e., $2.5\sigma_I$ above the mean threshold, where σ_I reflects largely the inter-subject variability). The RG thresholds in deutans (green circles, n=268) and protans (red circles, n=132) extend up to the upper limits of the display gamut for the parameters used in the CAD test (~ 34 standard normal CAD units). The least affected protans have thresholds around 4 CAD units, whilst the least affected deutans approach the UNL (~ 2 CAD units depending on age). The YB color thresholds in normal trichromats are similar to those shown above, but the thresholds increase faster with age, at 1.6% / year[1]

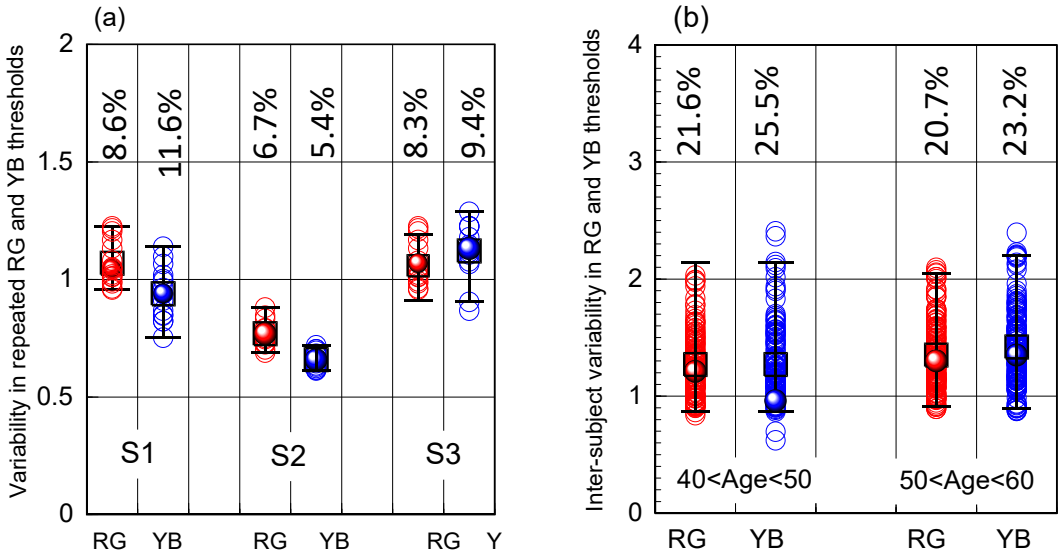
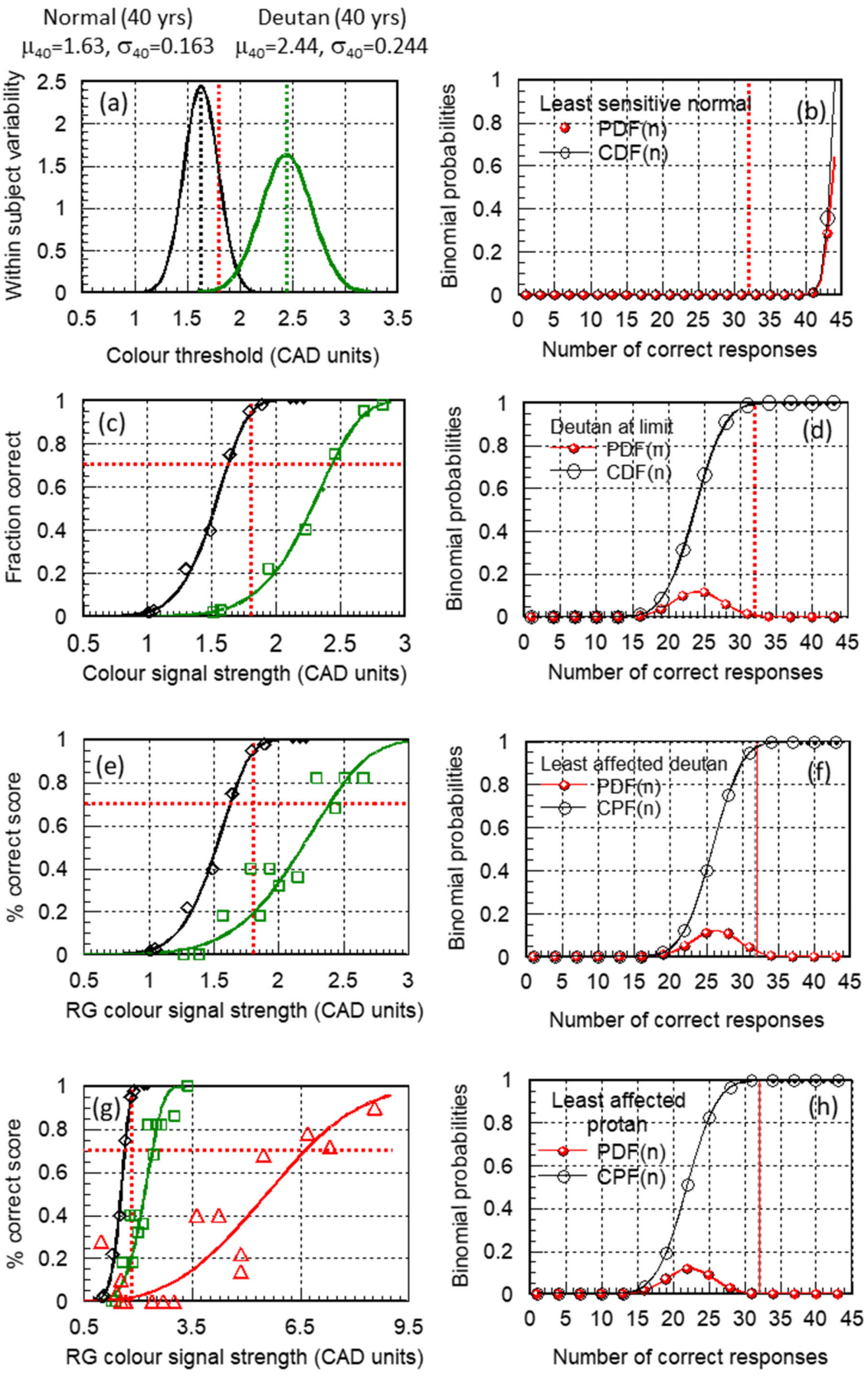


Fig. 4. Examples of within- (σ_w) and inter-subject variability (σ_I) when using the CAD test. Estimates of σ_w are based on 16 repeated threshold measurements of RG and YB chromatic sensitivity in three normal trichromats measured in this study. The S1 (19 years old male), S2 (42 years old female) and S3 (21 years old female). The inter-subject variability was estimated in two groups of normal trichromats: 40 to 50 years old, n=126, and 50 to 60 years old, n=148. The percentages shown in the graphs represent the coefficient of variation ($100*\sigma/\mu$). These estimates are based on a data set obtained in a previous study[1] (b). Note the different limits used along the ordinate for clarity of illustration in (b). Mean values, μ, are shown as black squares and median values as filled dots. There are only small differences between these two parameters. The results show much larger inter-subject variability with σ_I values ~ 2.7 times greater than σ_w values for both RG and YB thresholds. It is also of interest to note that the inter-subject variabilities in RG and YB thresholds remain relatively unchanged in the two age groups.


Normal (40 yrs)
μ_{40}=1.63, σ_{40}=0.163
Deutan (40 yrs)
μ_{40}=2.44, σ_{40}=0.244
(a)
Within subject variability
Colour threshold (CAD units)
(b)
Least sensitive normal
PDF(n)
CDF(n)
Binomial probabilities
Number of correct responses
(c)
Fraction correct
Colour signal strength (CAD units)
(d)
Deutan at limit
PDF(n)
CDF(n)
Binomial probabilities
Number of correct responses
(e)
% correct score
RG colour signal strength (CAD units)
(f)
Least affected deutan
PDF(n)
CPF(n)
Binomial probabilities
Number of correct responses
(g)
% correct score
RG colour signal strength (CAD units)
(h)
Least affected protan
PDF(n)
CPF(n)
Binomial probabilities
Number of correct responses

557

Fig. 5. Measured data and model predictions of sensitivity and specificity in the CVS test. Section (a) shows the expected distributions of repeated RG CAD thresholds in a 'least-sensitive' 40 years old normal trichromat (black) and the least-affected deutan (green), when the within subject variability (σ_w) is assumed to be 10% of the mean threshold. This assumption also applies to the predicted probability of a correct response for the least-affected deutan shown in (c). The Gaussian functions describe the expected distribution of CAD thresholds which correspond to the color signal strength that yields 71% probability of a correct response (indicated by the dotted, horizontal, red like in (c)). This analysis applies to any other age, but the color signal strength employed in the CVS test (indicated by the dotted vertical lines) varies with age. The Weibull function fitted to the actual data points for a least-affected deutan measured in the study is shown in (e). Each fitted function is then corrected for the 0.5 chance probability of a correct response to reflect the outcome of the CVS test and then used to compute the probability of a correct response, p, for the color signal strength employed in the test (i.e., $\mu_I + 2\sigma_I$, see text). This p-value, the chance probability of a correct response (0.5) and the number of trials (44) are then used to compute the binomial predictions that describe the outcome of the CVS test for a least-sensitive normal trichromat (b), for the model predictions for a least-affected deutan (d), for the measured data in a least-affected deutan (f) and for a least-affected protan (h). The dotted, vertical, red lines in sections b, d, f and h show the minimum number of correct responses needed to pass the CVS test (see text for full explanations of the graphs).

Section A	CVS Validation Study		
Classification of CV Class based on CAD	Number assessed	% of subjects who PASS the CVS test	
		Red / Green	Yellow / Blue
Normal Trichromats	27	100.0%	100.0%
Deutan-like deficiency	41	0.0%	100.0%
Protan-like deficiency	14	0.0%	100.0%
Tritanopia	0	N/A	N/A
Acquired loss in subjects with congenital deficiency	2	0.0%	0.0%
Σ	84		

Section B	Repeatability Study: 24 repeated CVS tests / subject		
Classification of CV Class based on CAD	Number assessed	% of PASSES in repeated tests	
		Red / Green	Yellow / Blue
Normal Trichromats	6	100.0%	99.3%
Deutan-like deficiency	3	0.0%	100.0%
Protan-like deficiency	0	N/A	N/A
Tritanopia	0	N/A	N/A
Acquired loss in subjects with congenital deficiency	0	N/A	N/A
Σ	9		

Table 1. Preliminary test results in 84 subjects investigated as part of the multi-center CVS validation trial started in December 2019. The trial will continue when the Covid-19 restrictions on face to face clinical assessment are removed. Every subject repeated the CVS test 24 times, for both RG and YB screening. All normal trichromats pass the CVS for RG assessment in every trial. One normal trichromat fails the YB assessment in one, out of 24 trials. All deutan subjects fail RG and pass YB in every trial.

Class of Colour Vision	p-value for RG_{CVS} = $\mu_{age}+2\sigma_I$	P ≥ 33 (Pass)	P ≤ 32 (Fail)
Least sensitive normal	0.971	1	0
Least affected deutan	0.597	0.025	0.975
Least affected protan	0.509	0.001	0.999

Table 2. CVS test outcomes based on a RG color signal strength $2\sigma_I$ above the mean, age-matched, CAD threshold. In order to pass the CVS test, one requires a correct response score ≥ 33. Note that the p-values of a correct response shown above are based on measured data for the least-sensitive normal trichromat and for the least-affected deutan and protan subjects who are the ones most likely to be misclassified. Since less than 5% of subjects within each category can be described as 'least-sensitive' normal trichromats and 'least-affected' deutans, the number of deutans that can potentially be misclassified is very small.

References

1. Barbur JL, Rodriguez-Carmona M. *Color vision changes in normal aging.* In: A.J. E, M.D. F, A. F (eds), Handbook of Color Psychology. Cambridge, UK: Cambridge University Press; 2015:180-196.
2. Rodriguez-Carmona M, Barbur JL. *Variability in normal and defective colour vision: Consequences for occupational environments.* In: Best J (ed), Colour Design (Second Edition): Woodhead Publishing; 2017:43-97.
3. Verriest G, Laethem Jv, Uvijls A. *A new assessment of the normal ranges of the Farnsworth-Munsell 100-hue test scores.* American Journal of Ophthalmology 1982;93:635-642.
4. Kinnear PR, Sahraie A. *New Farnsworth-Munsell 100 hue test norms of normal observers for each year of age 5-22 and for age decades 30-70.* Br J Ophthalmol 2002;86:1408-1411.
5. Paramei GV, Oakley B. *Variation of color discrimination across the life span.* J Opt Soc Am A Opt Image Sci Vis 2014;31:A375-A384.
6. Knoblauch K, Vital-Durand F, Barbur JL. *Variation of chromatic sensitivity across the life span.* Vision Res 2001;41:23-36.
7. Birch J. *Worldwide prevalence of red-green color deficiency.* J Opt Soc Am A Opt Image Sci Vis 2012;29:313-320.

8. Wright WD. *Researches on normal and defective colour vision.* London: Henry Kimpton; 1946.

9. Davidoff C, Neitz M, Neitz J. *Genetic Testing as a New Standard for Clinical Diagnosis of Color Vision Deficiencies.* Transl Vis Sci Technol 2016;5:2.

10. Pokorny J, Smith VC, Verriest G, Pinckers AJLG. *Congenital and Acquired Color Vision Defects.* New York, N.Y: Grune and Stratton; 1979:1-409.

11. Verriest G. *Acquired Colour Vision Deficiencies*. 1972.

12. Costa MF, Oliveira AG, Feitosa-Santana C, Zatz M, Ventura DF. *Red-green color vision impairment in Duchenne muscular dystrophy*. Am J Hum Genet 2007;80:1064-1075.

13. Pedersen HR, Hagen LA, Landsend ECS, et al. *Color Vision in Aniridia.* Invest Ophthalmol Vis Sci 2018;59:2142-2152.

14. O'Neill-Biba M, Sivaprasad S, Rodriguez-Carmona M, Wolf JE, Barbur JL. *Loss of chromatic sensitivity in AMD and diabetes: a comparative study.* Ophthalmic Physiol Opt 2010;30:705-716.

15. Abdel-Hay A, Sivaprasad S, Subramanian A, Barbur JL. *Acuity and colour vision changes post intravitreal dexamethasone implant injection in patients with diabetic macular oedema.* PLoS One 2018;13:e0199693.

16. Martin DR, Fowlkes CC, Malik J. *Learning to detect natural image boundaries using local brightness, color, and texture cues.* IEEE Trans Pattern Anal Mach Intell 2004;26:530-549.

17. Sumner P, Mollon JD. *Chromaticity as a signal of ripeness in fruits taken by primates.* J Exp Biol 2000 Jul ;203 Pt 13:1987 -2000 2000;203 Pt 13:1987-2000.

18. Barbur JL, Rodriguez-Carmona, M., Hickey, J., Evans, S., Chorley, A. *Analysis of European Colour Vision Certification Requirements for Air Traffic Control Officers.* London, UK: CAA (UK) Report, CAP 1429; 2016:1-76.

19. Dain SJ, Casolin A, Long J, Hilmi MR. *Color Vision and the Railways: Part 1. The Railway LED Lantern Test.* Optom Vis Sci 2015;92:138-146.

20. Dain SJ, Casolin A, Long J. *Color Vision and the Railways: Part 2. Comparison of the CN Lantern Used on the Canadian Railways and Railway LED Lantern Tests.* Optom Vis Sci 2015;92:147-151.

21. Cole BL. *Assessment of inherited colour vision defects in clinical practice*. Clin Exp Optom 2007;90:157-175.

22. Rodriguez-Carmona M, Evans B, Barbur JL. *Assessment of chromatic sensitivity: 2. Color assessment outcomes using single and multi-test protocols* Color Res Appl Nov 2020; DOI: 10.1002/col.22598

23. Barbur JL, Rodriguez-Carmona M. *Colour vision requirements in visually demanding occupations.* Br Med Bull 2017;122:51-77.

24. DeMarco P, Pokorny J, Smith VC. *Full-spectrum cone sensitivity functions for X-chromosome-linked anomalous trichromats.* J Opt Soc Am A 1992;9:1465-1476.

25. Neitz J, Neitz M. *The genetics of normal and defective color vision.* Vision Res 2011;51:633-651.

643 26. Neitz J, Neitz M, He JC, Shevell SK. *Trichromatic color vision with only two spectrally distinct*
644 *photopigments.* Nat Neurosci 1999;2:884-888.

645 27. Konstantakopoulou E, Rodriguez-Carmona M, Barbur JL. *Processing of color signals in female*
646 *carriers of color vision deficiency.* J Vis 2012;12.

647 28. Neitz M, Carroll J, Renner A, Knau H, Werner JS, Neitz J. *Variety of genotypes in males diagnosed*
648 *as dichromatic on a conventional clinical anomaloscope*. Vis Neurosci 2004;21:205-216.

649 29. Barbur JL, Connolly DM. *Effects of hypoxia on color vision with emphasis on the mesopic range.*
650 Expert Rev Ophthalmol 2011;6:409-420.

651 30. Barbur JL, Rodriguez-Carmona M, Harlow JA, Mancuso K, Neitz J, Neitz M. *A study of unusual*
652 *Rayleigh matches in deutan deficiency*. Vis Neurosci 2008;25:507-516.

653 31. Evans B, Rodriguez-Carmona M, Barbur JL. Assessment of chromatic sensitivity: 1. *Visual signals*
654 *that affect the outcome of color screening tests with emphasis on the Farnsworth D-15.* Color Res
655 Appl Nov 2020; DOI: 10.1002/col.22596.

656 32. Dain SJ, Atchison DA, Hovis JK. *Limitations and Precautions in the Use of the Farnsworth-Munsell*
657 *Dichotomous D-15 Test.* Optom Vis Sci 2019;96:695-705.

658 33. Squire TJ, Rodriguez-Carmona M, Evans AD, Barbur JL. *Color vision tests for aviation: comparison*
659 *of the anomaloscope and three lantern types*. Aviat Space Environ Med 2005;76:421-429.

660 34. Rodriguez-Carmona M, O'Neill-Biba M, Barbur JL. *Assessing the severity of color vision loss with*
661 *implications for aviation and other occupational environments*. Aviat Space Environ Med
662 2012;83:19-29.

663 35. Koefoed VF, Miles T, Cason JB, Troche R. *Colour vision classification - comparing CAD and CIE*
664 *143:2001 International recommendations for colour vision requirements in transport*. Acta
665 Ophthalmol 2020. DOI 10.1111/aos.14450

666 36. Fernandes TP, Santos NA, Paramei GV. *Cambridge Colour Test: reproducibility in normal*
667 *trichromats.* J Opt Soc Am A Opt Image Sci Vis 2020;37:A70-A80.

668 37. Regan BC, Reffin JP, Mollon JD. *Luminance noise and the rapid determination of discrimination*
669 *ellipses in colour deficiency*. Vision Res 1994;34:1279-1299.

670 38. Rodriguez-Carmona M, Barbur J. *Variability in normal and defective colour vision: consequences for*
671 *occupational environments*. Colour Design: Woodhead Publishing; 2017:43-97.

672 39. Barbur JL. *'Double-blindsight' revealed through the processing of color and luminance contrast*
673 *defined motion signals.* Prog Brain Res 2004;144:243-259.

674 40. CIE. Proceedings 1931. In: CIE (ed): Cambridge University Press; 1931.

675 41. Barbur JL, Harlow JA, Plant GT. *Insights into the different exploits of colour in the visual cortex.*
676 Proc R Soc Lond B 1994;258:327-334.

677 42. Chorley AC. *A modified protocol for color vision screening using Ishihara.* Aerospace Medicine and
678 Human Performance 2015;86 747-751.

43. Levitt H. *Transformed up-down methods in psychoacoustics.* Journal of the Acoustical Society of America 1971;49:467-477.

44. Jennings BJ, Barbur JL. *Colour detection thresholds as a function of chromatic adaptation and light level.* Ophthalmic Physiol Opt 2010;30:560-567.

45. Barbur JL, Evans B, Rodriguez-Carmona M. *New screener for congenital and acquired colour deficiencies*. Investigative Ophthalmology & Visual Science 2019;60:1314-1314.

Author biographies

John Barbur combines fundamental vision science with applied and clinical research. His main interests include mesopic vision, pupil responses and eye movements, blindsight, spatial and chromatic vision, scattered light and the development of novel techniques and instrumentation to study vision, both in the lab and in the clinic. Work on camouflage led to insights into the processing of luminance and color signals, the development of the Color Assessment and Diagnosis (CAD) test and more recently, the Color Vision Screener (CVS) test. These and other Advanced Vision and Optometric Tests (AVOT) were developed over several years and found important applications in both clinical work and in visually-demanding occupations.

Marisa Rodriguez-Carmona studied Physics at Imperial College London before receiving her doctoral degree for work on 'Variability in chromatic sensitivity' at the Centre for Applied Vision Research at City, University of London. She is currently a lecturer in Optics and Visual Science at City, University of London and leads the Physics and Optics modules. Her research interests include the study of sex related differences in color vision and the optimisation of color assessment protocols in the clinic and in occupations.

Benjamin Evans is a PhD student in the department of optometry and vision science at City, University of London. His research interests include color vision assessment and the development of novel methods for the efficient and effective screening of normal trichromatic color vision.