**Color Vision Assessment -** **2. Color assessment outcomes using single and multi-test protocols**

Marisa Rodriguez-Carmona

Benjamin E W Evans

John L Barbur

Affiliation:

Centre for Applied Vision Research, School of Health Sciences,

City, University of London, UK.

Corresponding author:

Marisa Rodriguez-Carmona

Email: marisa@city.ac.uk

Phone: +44 207 040 0262

ORCID:  [0000-0002-7366-6159](http://orcid.org/0000-0002-7366-6159)

**ABSTRACT:**

The main purpose of this study was to produce reliable, color assessment outcomes to examine the extent to which single- and multi-test protocols in use meet current clinical and occupational needs. The latter include the detection of small changes in chromatic sensitivity as the earliest signs of retinal and / or systemic disease, and the need to assess the class of color vision in congenital deficiency and to quantify severity of loss.

Color vision was assessed using Ishihara (IH), Farnsworth Munsell D-15, City University (CU, 2nd ed.) and Holmes-Wright type A (HW-A) lantern tests. All subjects also carried out Colour Assessment and Diagnosis (CAD) and Nagel anomaloscope tests. The sample included 350 normal trichromats, 1012 deutans and 465 protans (age 31.1 ± 12.4, range 10-65 years).

The results reveal the trade-off between sensitivity and specificity, depending on the number of errors accepted as a pass on the IH test. The D-15 and CU tests pass all normals and almost 50% of subjects with color vision deficiency. The HW-A lantern passes all normals, 22% of deutans and 1% of protans. The multi-test protocols designed to identify protans and to pass only subjects with mild color loss, pass over 50% of protans and deutans. Many of the subjects who fail exhibit less severe loss of color vision than others who pass. When high sensitivity for detection of congenital deficiency is achieved, single-test protocols fail many normal trichromats. Multi-test protocols produce large variability and fail to achieve desired aims.

**Keywords:** color tests, congenital color deficiency, acquired color deficiency, red-green color vision, Ishihara, Farnsworth Munsell D-15, City University (2nd ed.), CAD test, color assessment protocols

**1. INTRODUCTION**

The visual system relies on at least four mechanisms that evolved to extract edges and contours and to segment areas of interest based on ‘luminance’ and ‘color’ signals1, 2. These mechanisms rely on the normal functioning of rods and short (S), middle (M) and long (L) wavelength sensitive cones3. The perceptual representation of objects the eye can see relies on the strengths of these signals. Color signals contribute significantly to what we see and play an important part in enhancing our visual performance4-8. Colored objects are more conspicuous and they are also processed in parallel over the visual field bringing considerable advantages to visual search8-10. Color coding is used in signal lights to communicate useful information, to segment regions of interest within a larger scene and to group together spatially discrete objects into meaningful subgroups, all contributing to the enhancement of visual performance11, 12. These advantages are best achieved with normal trichromatic color vision which relies on a unique set of cone pigment optical densities and spectral responsivity functions with only small inter-subject variation13, 14. When one or more of these photoreceptor pigments are absent or simply replaced by variant pigments, colors appear to have different hues, are less saturated or even absent, and side by side metameric colors no longer match15. The most common color deficiency is caused by genetic inheritance of variant pigments in either L or M cones16. Subjects with protan deficiency rely on normal and variant M cone pigments and are labelled protanomalous. In the absence of a variant M cone pigment, the subjects are known as protanopes. Subjects with deutan deficiency that rely on normal and variant L cone pigments are known as deuteranomalous. Those with only a single L-cone pigment are known as deuteranopes17. Neither the deuteranopes nor the protanopes have red/green (RG) color vision, but they have different spectral luminance efficiency functions, with the protanopes exhibiting significant loss of sensitivity over the long wavelength region of the visual spectrum18. Some small spectral differences between cone photoreceptors arise as a result of significant differences in pigment optical density, even when only one distinct photopigment is involved. Such differences can result in small, residual RG color discrimination ability19. The absence of functioning S cones is a rare genetic condition and causes tritanopia with complete loss of yellow/blue (YB) color vision. What is more common is the loss of YB chromatic sensitivity in patients with diseases of the retina and the optic nerve20-22. Many of these patients also develop loss of RG color vision as a result of damage to cone photoreceptors or / and the neural pathways involved in the processing of color signals. Acquired color deficiency is more common in older subjects and very common in patients with diabetes and glaucoma23-26. Screening for congenital or / and acquired color deficiency in the clinic and the need to quantify severity of loss in order to produce reliable minimum color vision requirements within visually demanding occupations has encouraged the development of novel tests which isolate fully the use of RG and YB color signals and quantify more reliably the severity of color vision loss2, 27-30. The use of such tests remains limited and current, color assessment protocols rely mostly on conventional tests to assess the applicant’s class of color vision and severity of loss. Commonly used tests include the Ishihara (IH) pseudoisochromatic test plates, color arrangement tests such as the Farnsworth Munsell D-15 and City University (CU) and lantern tests such as the Holmes-Wright type-A lantern (HW-A) employed in the UK.

The IH test(Kanehara & Co. Ltd., Tokyo, Japan) is the most widely accepted screening test for congenital protan and deutan defects. Since its publication in 1917 it has been reprinted in various editions, including a full, abbreviated and concise version which contain 38, 24 and 14 plates respectively31. The abbreviated and concise versions are subsets of the full version. The latter has been recommended for routine clinical use31. Both the D-1532 and CU33 tests screen for ‘moderate and severe’ color vision loss34, 35. The D-15 panel contains 15 movable color samples judged to have approximately equal hue steps when illuminated with daylight. The CU (2nd ed. ) test consists of a series of ten plates. Each plate displays five colored discs selected from the Munsell series mounted on a black matt background: one central and four peripheral colors of equal size. The first six plates contain large circles and the last four smaller circles. The HW-A lantern test36 presents one of nine pairs of vertical color combinations37; two reds, two greens and one white. The lights have (x, y) - chromaticity co-ordinates within the limits for signal lights agreed internationally by the CIE38. The IH, D-15 and CU tests are commonly used in clinics and in occupational environments, whilst the HW-A is mainly an occupational test.

Protocols based on multiple tests have been produced to screen for normal trichromatic color vision as required in the most visually demanding occupations (such as air traffic control) or to pass all normal trichromats and color deficient subjects with mild loss of RG chromatic sensitivity (as practiced in the electrotechnical industry). Protocols have also been designed to pass subjects with a pre-determined level of RG color vision loss and to exclude all protans (e.g., UK fire fighters and special duties police officers). The efficiency of these protocols have not been examined thoroughly before, largely because conventional color assessment tests often fail to isolate RG and YB color signals and to diagnose accurately the applicant’s type of color vision. Equally important, until very recently it has not been possible to assess accurately the strength of RG color loss in applicants who pass and those who fail a given protocol. Furthermore, YB deficiency and the presence of acquired loss of color vision are not normally tested for, even when the job requirement specifies normal ‘trichromatic’ color vision.

This manuscript is the second in a series of three papers which examines the most common color evaluation tests and protocols employed in the clinic and also in visually-demanding occupations in relation to the type of color vision deficiency and the severity of loss in subjects who pass and fail each protocol. The first paper examines the visual signals subjects can utilize to pass the Farnsworth D-15 test. The purpose of this study (paper 2) was to examine how well single- and multi-test color assessment protocols achieve their objectives. Not least, the study also aimed to produce reliable statistics on the color assessment outcomes of common protocols to establish how well these protocols serve current needs. The third paper examines the fundamental limits of what one can achieve in color vision assessment and describes a new color vision screener test that approaches these limits.

**2. METHODS**

**2.1 Single color assessment tests**

In total, we investigated 1827 subjects with the IH test, Nagel anomaloscope (type I) and Colour Assessment and Diagnosis (CAD) test. There was 100% agreement between the classification made by the CAD test and the Nagel for all subjects in this study. From this sample, 674 subjects were also examined on the D-15 test, 636 subjects on the CU test (2nd ed.) and 359 subjects on the HW-A lantern test.

The reading errors on the first 25, or first 21 plates in the full version, or the first 17, or first 15 in the abbreviated version containing single- or double-digit numbers were examined. All subjects read plate 1 correctly. The most common pass criteria using the IH test are summarized in Table 1. The D-15 test results were evaluated using the ‘visual inspection’ method that has been described previously39. The number of major crossings and adjacent transpositions were recorded and the classification of the D-15 (the number of ‘protan/deutan/tritan’ crossings) was determined using the method described in the first paper in this series40. The subject’s task on the CU test is to select which of the four peripheral colors is perceived to be least different to the central color. Three peripheral colors represent typical isochromatic confusions with the central color, for each type of defect. The fourth color is an adjacent color in the D-15 sequence and is designated as the normal choice41. The number of errors on the 10 plates, as well as errors made on the ‘large’ circle plates, were recorded. The IH plates, the D-15 test caps and the CU plates were all illuminated with a Macbeth easel lamp (which approximates daylight at 6500K).

The results of both the Nagel and CAD tests were used to diagnose the type of color vision deficiency and the latter was used to provide a measure of severity of color vision loss. The Nagel measures Rayleigh matches42 and can be used to distinguish accurately between deutan and protan deficiency and between anomalous trichromacy and dichromacy43. The anomaloscope is, however, less useful as a way of quantifying severity of color vision loss, i.e. the parameters of the match cannot be used reliably to quantify a subject’s overall loss of chromatic sensitivity18, 44, 45.The CAD testmeasures RG and YB color thresholds using a technique that provides effective masking of luminance contrast signals without affecting chromatic sensitivity28, 46. The test has been fully described in earlier publications2, 40. The RG and YB CAD thresholds are approximately linearly proportional to the cone contrasts generated by the colored stimulus47. This observation justifies why the measured thresholds are appropriate to quantify severity of color vision loss. The average healthy, young, normal trichromat has RG and YB CAD thresholds around 1 CAD unit. The maximum chromatic saturations that can be generated within the constraints of the CAD test depend on the characteristics of the stimulus display employed and can be as large as 37 CAD units for RG and 18 CAD units for YB discrimination2.

**2.2 Multi-test protocols**

Three multi-test protocols selected for examination in this study are shown in Table 1.

Protocol A is a stringent protocol which aims to pass only subjects with normal trichromatic color vision. As initial screening, this protocol employs either the first 25 plates of the 38-plate ed. or the first 15 plates of the 24 plates ed. of the IH test (Table 2). When the applicant fails the screener test, the HW-A lantern is used and is passed if no errors are made at 6m using the standard photopic protocol37.

Protocol B is less stringent and its objective is to pass all normal trichromats and also subjects with reduced chromatic sensitivity, but to exclude all those with protan deficiency. The protocol employs the IH (38 plates), with 4 or fewer errors on plates 1-21. If the applicant fails, two further secondary tests are administered and the applicant has to pass both tests. A pass on the D-15 requires fewer than two major crossings. A pass on the CU test (2nd ed.) requires four or fewer errors with no more than one error on the ‘large’ circle plates. In addition, an applicant diagnosed as having protan deficiency on 2 of the 3 tests carried out, also fails the protocol.

Protocol C is intended to fulfil the requirements of the fire service and aims to pass all normal normal trichromats and deutan subjects with mild colour vision deficiency and to exclude all protans. The protocol uses the IH (24 plates) with 2 or fewer errors on plates 1-17, followed by the D-15 with 2 of fewer adjacent transpositions needed to pass. If the applicant passes the D-15 test, an anomaloscope test is required to identify and exclude protans.

**2.3 Participants**

The study participants were mainly occupational applicants referred to the Advanced Vision and Optometric Tests (AVOT) service at City, University of London. Subjects carried out the color vision tests described above. Subjects with best corrected visual acuity less than 20/40, or clinical signs of pathology or ocular abnormalities were excluded from the study. The mean age was 31.1 ± 12.4 year with a median of 29 years and age range of 10-65 yrs. Informed consent was obtained from all subjects. The tenets of the Declaration of Helsinki were followed and the study complied with the City, University of London research and ethical guidelines.

**3. RESULTS**

**3.1 Single test protocols**

Table 2(a) shows the percentage of normals, deutans and protans who pass four, commonly used IH test protocols. Table 2(b) and 2(c) show similar results for the D-15 and the CU (2nd ed.) tests. The maximum CAD RG thresholds for normal, deutan and protan subjects who pass each protocol are shown in the gray columns. When no errors are allowed on the first 15 plates, 11% of normals fail and a very small number of color deficient subjects pass. When the pass criteria is ≤ 2 errors the number of color deficient subjects who pass increase to 3.8% deutans and ~1% protans. Some of the color deficients applicants who pass can have severe loss of RG color vision with thresholds as high as 13 RG CAD units. Table 2(b) reveals that a large number of protans and deutans pass the D-15 for two different pass criteria, some with RG thresholds close to 25 CAD units. When no errors are allowed on the CU test, almost all normal trichromats pass (99.3%), but ~46% of deutans and ~39% of protans also pass (Table 2c). When a maximum of 4 errors are allowed (with only 1 error on the large circle plates), 66% and 56%, of deutans and protans, respectively, are able to pass.

**3.2 Multi-test Protocols**

A summary of the pass rates for each multi-test protocol is shown in Table 3. Protocol A passes all normal trichromats, 22% of deutans and 1.1% of protans. The maximum RG thresholds of deutans and protans who pass the HW -type A lantern test are 12.0 and 9.4 CAD units, respectively. Figures 1 and 2 show the spread in the RG CAD thresholds of subjects that pass and fail protocol A. For protocol B, all normals, 55.6% of deutans and 46.8% of protans pass. The deutans and protans who pass can have severe loss of RG color vision with thresholds as high as 22 and 24 CAD units, respectively. Figures 3 and 4 show the results of subjects who pass and fail protocol B. Protocol C passes all normals, 54.8% of deutans and 1.8% of protans with maximum RG thresholds of 17.7 and 13.3, respectively. Figures 5 and 6 show the results of subjects who pass and fail protocol C.

**4. DISCUSSION**

**4.1 Single test protocols**

The results reveal how well each test and protocol accomplishes its objectives, whether these require applicants to have normal color vision, mild loss or to exclude all those with a specific type of deficiency. Table 2(a) shows that the IH test can achieve high sensitivity when screening for congenital deficiency, however this is at the cost of failing up to 19% of normal trichromats (38-plate ed.). The test isolates the use of RG color signals by using random luminance variation in the dots that make up each test plate. Due to its simplicity and availability, the IH is relatively easy to learn, making use of cues other than color to carry out the test48. In addition, the orientations of the YB color confusing axes in deutan and protan subjects differ significantly from each other and also from the orientation expected in normal trichromats. The latter cannot make much use of YB color signals and many fail the full IH test when no errors are allowed (Table 2a). The YB color signals generated by the same stimuli are expected to be different in deutan and protan subjects and these differences may be sufficient to enable some subjects to carry out the task involved49, 50. This and other factors may account for the large range of thresholds for those who pass and those who fail the IH, D-15 and CU 2nd ed tests40.

Since subjects with more severe RG loss tend to fail more IH plates, the number of errors the subject makes on the IH test has been incorrectly assumed to be a good measure of the severity of RG color vision loss. In practice, the number of errors subjects make correlate very poorly with severity of loss51. In spite of these shortcomings, which cause large variability amongst subjects with similar RG color vision loss2, the IH test remains the most sensitive screener for congenital RG color deficiency, provided one is willing to accept a large number of false positives.

The D-15 and CU tests are unsuitable because of the large number of congenital color deficients who pass, in spite of severe loss of RG color vision (see Table 1b, and 1c). It is of interest to examine why the D-15 and CU tests are still in current use, either as single tests or part of multi-test protocols. The justification for the continued use of these tests is based on the desire to pass color deficient subjects with mild to moderate loss of RG chromatic sensitivity and to facilitate the identification of protans who are considered unsafe to work in certain environments where protans may be disadvantaged52. The D-15 and CU tests, when used in isolation, fail to achieve either of these objectives40, 53.

**4.2 Multi-test protocols**

Table 3 lists the outcome of the three multi-test protocols investigated in this study. Protocol A was designed for use in aviation and other transport related occupations. Although the initial screening component fails 11-19% of normals (depending upon the version of IH test employed), secondary testing using the HW-A lantern ensures that all normal trichromats and 22% of deutans pass and are indistinguishable from the normals who pass. One protan subject with a RG threshold of 9.4 passes protocol A, with zero errors on the IH and HW-A lantern test (see Table 3). It is worth noting that protocol A performs much better than protocol B in passing less severe color deficient subjects and excluding protans. Figure 1 shows the large variability in RG color vision loss in the deutans who pass (range ~2 to 12 units) and those who fail (range ~2 to 28 units) protocol A. Figure 2 shows the same deutans ranked according to their RG CAD thresholds. When the outcome of the protocol is evaluated relative to the severity of RG loss, one finds a large overlap in RG thresholds between those who fail and those who pass. The protocol is therefore not dichotomizing the population on the basis of severity of RG loss (Figure 1). All deutans with a RG threshold below 2.35 CAD units pass the HW-A lantern test. Only ~ 6% of the least-affected deutan subjects fall into this category. The HW-A lantern shows considerable variability of outcome in repeated tests2. As a result of this large variability and also the smaller, but significant within-subject variability in CAD thresholds54, there is no clear distinction between deutans who fail and those who pass the HW-A lantern (Figure 1). The dashed black line in Figure 2 corresponds to an ‘equivalent’ threshold of 4 CAD units. In the absence of a detailed study designed to establish pass / fail limits specific to occupational environments that employed protocol A, an ‘equivalent’ threshold of 4 CAD units can act as an acceptable compromise with significant advantages. Passing all deutans with a RG threshold ≤ 4 CAD units results in diminished variability, but the percentage of deutans who pass remains unchanged. The number of deutans who fail the protocol with a RG threshold below 4 CAD units equals the number of deutans who pass with a threshold above 4 CAD units. Figures 1 and 2 show clearly that the upper threshold limit of those who pass protocol A can extend to 12 CAD units. A protocol based on an ‘equivalent’ threshold of 4 CAD units exhibits much reduced variability because of the much smaller within-subject variability in the CAD test and the absence of any other cues the subject could use to pass the test48, 54. A protocol based on an ‘equivalent’ CAD threshold enhances safety by excluding subjects with severe deficiency who pass and is more equitable to those less severe subjects who fail the current protocol. This statistical approach can be applied to any protocol employed in other occupations. Doing so yields significant improvements over current practices, but is not the ideal outcome one can achieve. Detailed studies carried out in the aviation environment to establish how pilots with normal color vision and those with deutan and protan deficiency carry out the safety-critical, color-related tasks produced new pass / fail limits specific for deutan and for protan subjects. As a result of such studies8, 55 some aviation authorities have replaced multi-test protocols consisting on IH followed by a secondary test with a simpler protocol based on IH test followed by CAD only in those who fail or the CAD test alone. Deutans with thresholds ≤ 6 and protans with thresholds ≤ 12 CAD units pass the CAD based protocol with advantages such as reduced variability and more equitable outcome. The CAD pass / fail limits are also specific for protan and deutans. As a result of adopting an evidence-based protocol, 35% of all color deficient applicants pass and are now able to work as pilots in commercial aviation. In occupations which require normal trichromatic color vision such as Air Traffic Control, it is also important to allow for the effects of normal aging in order to pass all normal trichromats.

Protocol B is the most complex in current use and is arguably the least effective of the three protocols in achieving its stated aims. The protocol aims to pass all normal trichromats whilst excluding those with either severe loss of RG color vision and / or those with a protan deficiency. The protocol employs IH (38-plate ed.) with ≤ 4 errors on plates 1 to 21 followed by D-15 and CU tests for all subjects who fail the IH test. All subjects diagnosed as protan in at least two of the three tests also fail the protocol (Table 1). In spite of its intricacy, the protocol produces poor results; 56% of deutans and 47% of protans pass, some with severe loss of RG color vision (Table 3). Figure 3 shows the spread in the severity of RG color vision loss of those who pass and those who fail protocol B. The results show the unjust potential outcomes of this protocol with many less severe deutan and protan applicants who fail and many others who pass with significantly more severe loss of RG chromatic sensitivity. When ranked according to their CAD thresholds, the deutans and protans who pass and fail show considerable overlap (Figure 4). ‘Equivalent’ CAD thresholds needed to pass the same percentage of deutans and protans as the current protocol B can also be computed using the statistical approach described for protocol A. The dotted lines in Figure 4 show the ‘equivalent’ thresholds of 13.9 and 17.7 CAD units needed to pass the same number of deutans and protans, respectively.

The summary shown in Table 3 and the results summarised in Figures 3 and 4 demonstrate clearly that protocol B fails to achieve its aims of excluding protans and of passing only applicants with mild to moderate RG color deficiency. The aims of protocol B can in fact be best met using protocol A since the latter eliminates all protans and passes only 22% of the less severe deutans.

Protocol C relies on the use of IH with ≤ 2 errors on the first 17 plates of the 24 plates ed. followed by D-15 and the anomaloscope in those who fail. The protocol excludes protans by using the anomaloscope test which requires an experienced clinician, except for the very few protans who pass the IH test with ≤ 2 errors (Table 3). The protocol produces significant overlap in the deutans who pass and those who fail (Figure 5). Protocol C excludes all protan subjects who fail the IH test. Two protans with RG thresholds of 13.3 and 9.4 CAD units pass the protocol C criteria for the IH screening test and, under the normal protocol conditions, would not be required to take the anomaloscope test. Protocol C has an ‘equivalent’ threshold of 12.8 CAD units (Figure 6).

The results presented in this manuscript show that current tests and protocols fail to achieve their stated aims for various reasons. None of the tests employed in the multi-test protocols account for the expected changes in chromatic sensitivity as a result of normal aging56. The tests fail to exclude the potential use of other cues and may also fail to achieve adequate isolation of only RG or only YB color signals. As a result, the tests do not achieve high sensitivity and specificity and fail to quantify reliably the extent of RG and YB loss of color vision. The CAD test isolates the use of RG and YB color signals using dynamic luminance contrast noise57 to mask the detection of luminance contrast signals and uses several color directions to allow for potential variation in the orientation of color confusion lines. The CAD test has relatively small within subject variability54 and offers the choice of carrying out an ‘enhanced’ threshold when the subject’s first test yields a value within two standard deviations of the expected pass / fail limit for the selected environment. This procedure halves the standard error of the measured threshold and this enhances the efficiency of the test. Equally important, the test quantifies severity of color vision loss based on thresholds which are directly proportional to the cone contrasts generated by the colored stimulus and classifies the subject’s type of color vision with the same accuracy that can be achieved using the anomaloscope. This is important in some occupations when it is desired to exclude protan subjects or to set differential pass / fail limits for deutan and protans, as has been done in commercial aviation2. The high sensitivity of the CAD test and assessment of both RG and YB thresholds makes the test particularly useful in the clinic. The disadvantages of the CAD test include the high cost of fully calibrated equipment and the relatively long time needed for full RG and YB color assessment (~ 12 to 15 minutes). A new, ‘two-test’ protocol based on the initial use of a rapid Colour Vision Screener with close to 100% sensitivity reduces to 6% the number of applicants required to take the full CAD test. This estimate assumes equal numbers of males and females and 2% prevalence of acquired deficiency54.

In the absence of minimum color vision requirements that can be classed as safe within a given occupation, the CAD test makes it possible to use an ‘equivalent’ threshold to reduce the variability associated with current protocols without affecting the number of deutans and protans who pass.

Technological advances in visual displays and improved understanding of the limits of color assessment in human vision can be used to overcome many of the limitations of conventional color assessment tests and lead to safer and fairer color assessment protocols for use both in occupations and in the clinic54.

**5. CONCLUSIONS**

Single and multi-test protocols based on conventional color tests fail to meet current color assessment requirements. The IH test has the highest sensitivity for detection of subjects with RG congenital deficiency, but the test has poor specificity and fails to identify accurately the subject’s type of color deficiency and severity of loss.

The D-15 and the CU (2nd ed.) tests are inappropriate for occupational use since they fail to diagnose reliably the subject’s type of color deficiency and to estimate the severity of color vision loss.

The use of secondary tests with equally poor sensitivity and specificity fails to solve the problem, particularly when the aim of the protocol is to exclude protan applicants and to pass only those with color vision loss below a specified level.

Protocols based on the ‘equivalent’ threshold limits reported in this manuscript reduce variability, enhance safety by excluding subjects with severe deficiency and are more equitable to those less severe subjects who fail current protocols. Equally important, the number of color deficient subjects expected to pass ‘equivalent’ threshold protocols remains unchanged.

**Acknowledgements**

We acknowledge the UK Civil Aviation Authority, the Colt Foundation and City, University of London for financial support with this study.

**Disclosures:**

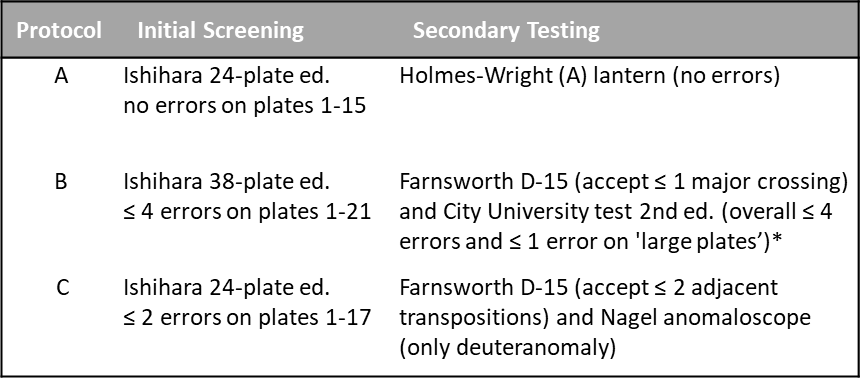
**Marisa Rodriguez-Carmona**, City, University of London, None.

**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**

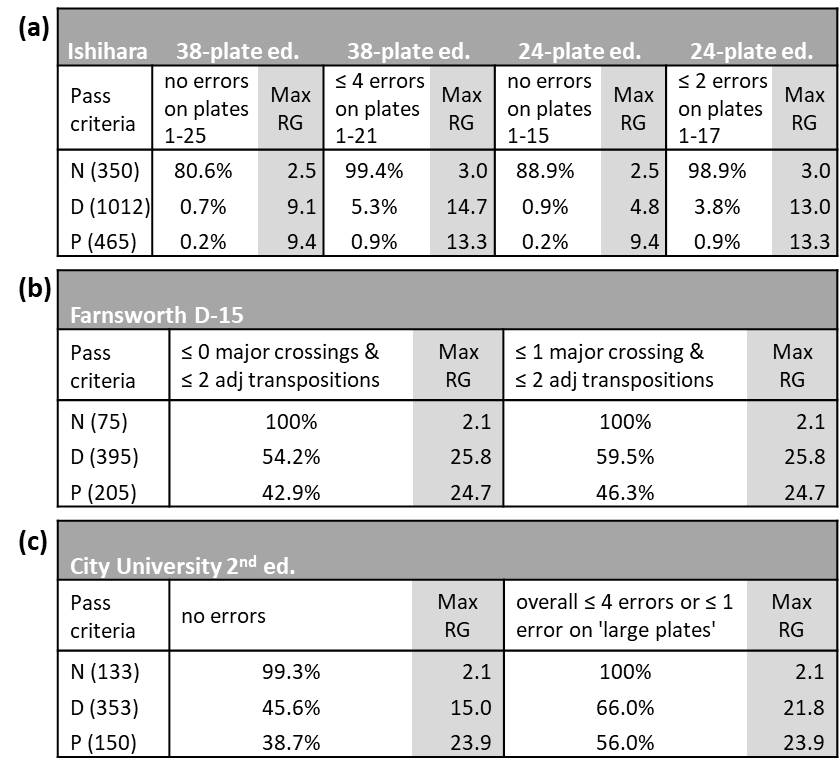
**Multi-test color assessment protocols employed in occupational settings**

****

\*Protocol B also excludes all applicants diagnosed with protan deficiency based on 2/3 of the tests employed.

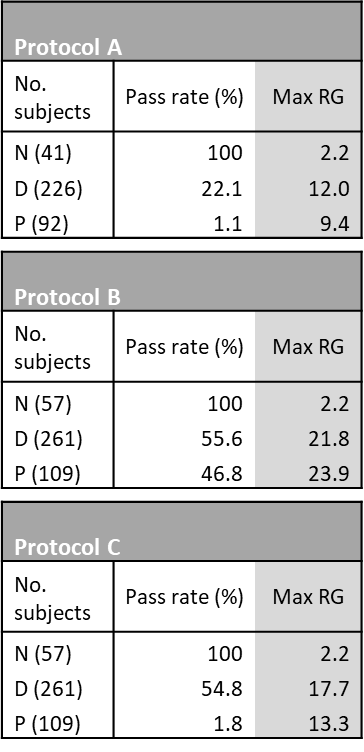
**Table 1.** Summary of three multi-test protocols developed to assess minimum color vision requirements in visually demanding and safety-critical occupations such as aviation, police and fire services.

**Outcome of single-test protocols**



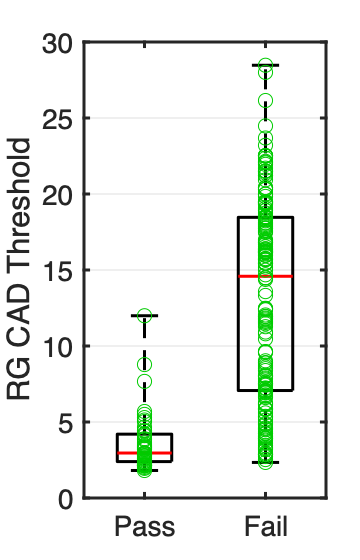
**Table 2.** The statistical outcomes of four, commonly used protocols based on the Ishihara test are shown in **(a).** The statistical outcome of two common pass criteria used on the Farnsworth D-15 and City University (2nd ed., CU) tests are shown in **(b)** and **(c)**, respectively. The percentage of subjects who pass each protocol is shown for each class of color vision (classification based on Nagel and CAD test) together with the maximum RG CAD threshold observed within each class. The upper RG limit for young, normal trichromats is ~ 1.85 CAD units. The results in (a) are based on 1827 subjects (350 normals (N), 1012 deutans (D) and 465 protans (P)) and show that when no errors are allowed on the first 25 plates of the 38 plates ed., the test fails ~ 19% of normal trichromats and almost all protans and deutans. (b) is based on 675 subjects (N=75, D=395 and P=205) and reveals the large percentages of deutan and protans who pass the D-15 test, some with RG color thresholds approaching 25 CAD units. (c) is based on 636 subjects (N=133, D=353 and P=150) and shows the percentage of subjects within each class who pass the CU test. When no errors are allowed, almost all normal trichromats pass, but ~46% of deutans and ~39% of protans also pass. Protocols based on this test often allow 4 or fewer errors on the small plates and a maximum of one error on the large plates. The net result is to increase further the percentage of deutans and protans who pass the test, some with severe loss of color vision.

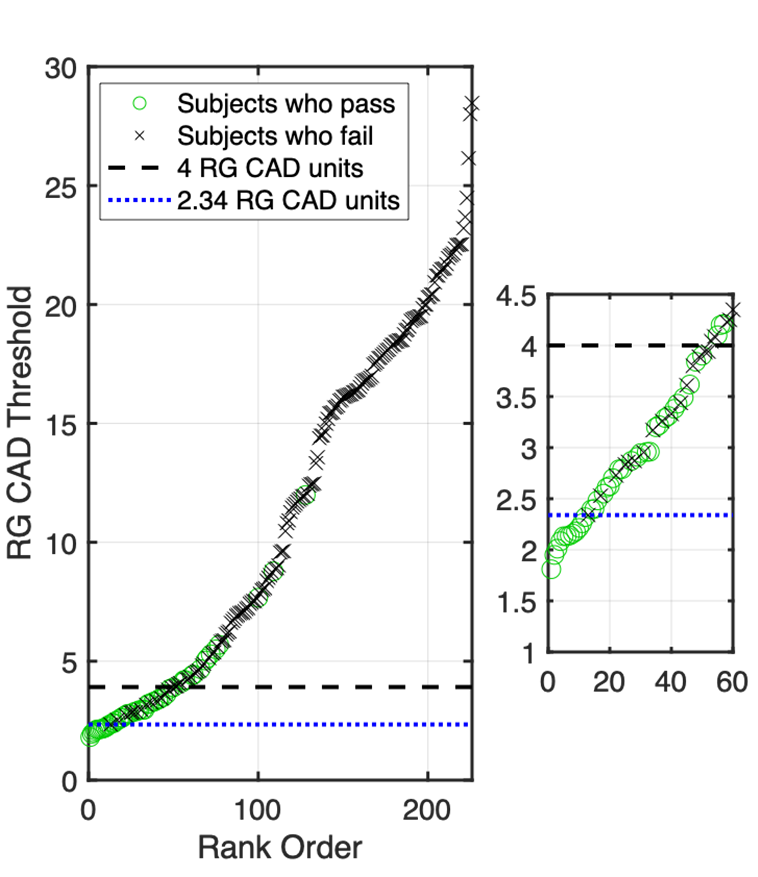
**Outcome of multi-test protocols**



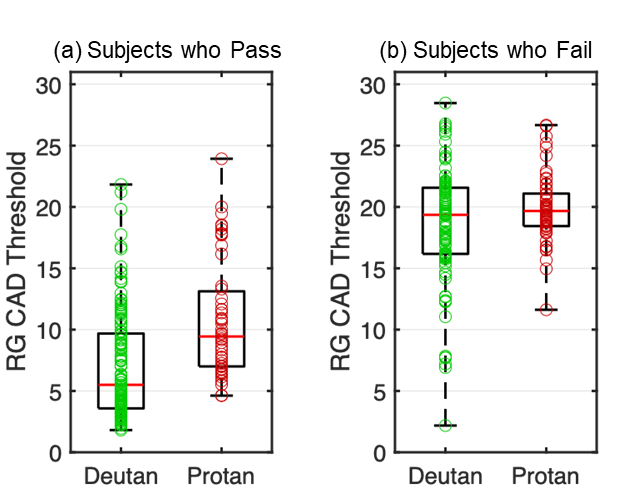
**Table 3.** Statistical outcomes of protocols A, B and C showing the percentage of applicants within each color vision class (N for normals, D for deutans and P for protans, based on Nagel and CAD test classification) who pass each protocol. The ‘Max RG’ column lists the highest RG CAD color threshold within each class. The upper RG limit for young, normal trichromats is ~ 1.85 CAD units. Protocol A, based on Ishihara 24 (or 38) plates ed. with zero errors followed by Holmes-Wright type A lantern test passes all normal trichromats, ~22% of deutans and ~1% of protans. The outcome of protocol A is identical to the HW-A lantern when used on its own. Protocol B, based on Ishihara followed by Farnsworth D-15 and City University (2nd ed.) tests passes all normal trichromats, ~56% of deutans and ~47% of protans. The protocol passes more than half of all subjects with congenital RG deficiency. Protocol C is based on Ishihara (with two or fewer errors on the first 17 plates of the 24-plates ed.), followed by Farnsworth D-15 (with 2 or fewer major crossings accepted as a pass). Those who pass are then further tested on the Nagel anomaloscope to ensure the applicant is deuteranomalous. All normals trichromats, ~55% of deutans and just below 2% of protans pass this protocol.

**Figure 1.** The distribution of red-green (RG) CAD thresholds of deutans who pass and fail protocol A. The results are from 226 deutan subjects examined with the HW-A lantern test. The median RG CAD threshold for deutans who pass (2.96 CAD units) is smaller than the median value for deutans who fail (14.59 CAD units) the HW-A lantern test. The boxplots show IQR (25%-75%) and the median threshold (red line).

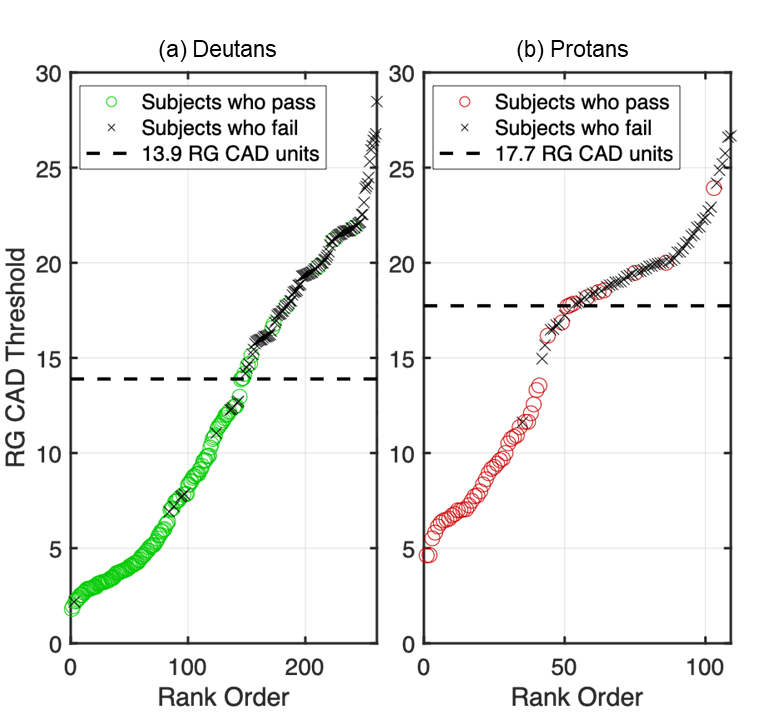




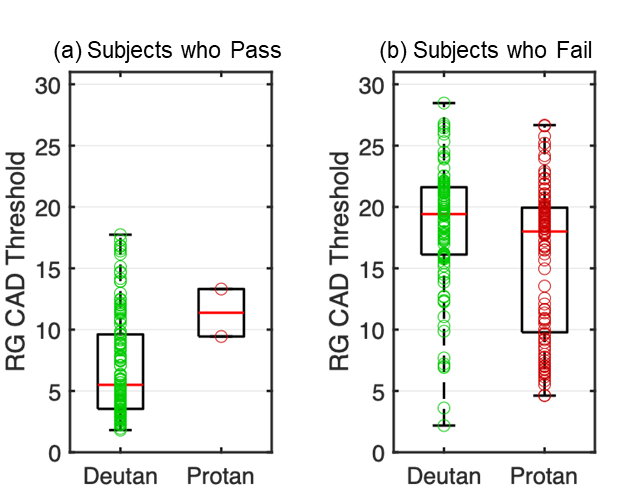
**Figure 2.** Ranked RG CAD thresholds for deutans subjects who pass and fail protocol A. The results for protans are not shown as almost all fail the HW-A lantern. The ranking of RG CAD thresholds for deutan subjects and the labelling of those who pass and those who fail the HW-A lantern reveal the significant overlap in RG CAD thresholds among deutans who pass and deutans who fail the lantern test. This is due mainly to the large within subject variability observed in repeated HW-A lantern tests58 and the much smaller within subject variability observed in CAD thresholds54. The results also illustrate clearly the inequity of this protocol when less severe deuteranomalous subjects fail and other subjects with more severe loss pass. All deutans with thresholds less than ≤ 2.34 CAD units pass the HW-A test. The inset on the right is a magnified region of the main graph added to illustrate this finding. Only ~6% of the least affected deutans fall into this category.



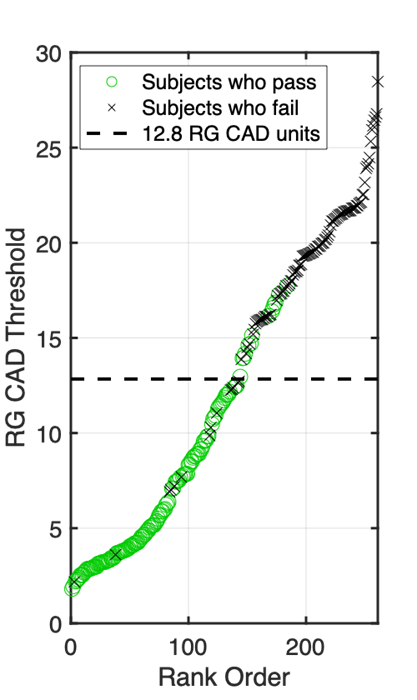
**Figure 3.** The distribution of red-green (RG) CAD thresholds in 427 color deficient subjects who pass (a) and fail (b) protocol B (see Tables 2 & 3). All normal subjects (N=57) passed (not shown), 196 subjects (145 deutans and 51 protans) pass, whilst 174 subjects (116 deutans and 58 protans) fail protocol B. The median RG CAD threshold in the subjects who pass (5.50 and 9.44 CAD units, for deutans and protans, respectively) is smaller than the corresponding median value in those subjects who fail (19.36 and 19.67 CAD units, for deutans and protans, respectively). Boxplots show IQR (25%-75%), median (red line). The whiskers represent the minimum and maximum RG CAD thresholds.



**Figure 4.** Ranked red-green (RG) CAD thresholds for deutan (a) and protan (b) applicants who pass and fail protocol B (Table 2). The equivalent RG threshold for deutans is 13.9 and for protans 17.7 CAD units. This threshold is established statistically and represents the value in RG CAD units for which the number of subjects who fail protocol B with thresholds lower than this threshold value is equal to the number of subjects who pass protocol B with thresholds higher than this value.



**Figure 5.** The distribution of red-green (RG) CAD thresholds in 427 color deficient subjects who pass (a) and fail (b) protocol C (see Tables 2 & 3). All normal subjects (N=57) passed (not shown), 145 subjects (143 deutans and 2 protans) pass, whilst 225 subjects (118 deutans and 107 protans) fail protocol C. The median RG CAD threshold for subjects who pass (5.50 and 11.38 CAD units, for deutans and protans, respectively) is smaller than the median value for those subjects who fail (19.42 and 18.09 CAD units, for deutans and protans, respectively). The two protans 13.31 and 9.44 CAD units who pass the protocol, passed the initial screening with the Ishihara test. Boxplots show IQR (25%-75%), median (red line). The whiskers represent the minimum and maximum in the spread of all the data.



**Figure 6.** Ranked red-green (RG) CAD thresholds for deutan applicants who pass and fail protocol C (Table 2). The ‘equivalent’ RG threshold for deutans is 12.8. This threshold is established statistically and represents the value in RG CAD units for which the number of subjects who fail protocol C with thresholds lower than this threshold value is equal to the number of subjects who pass protocol C with thresholds higher than this value. The protocol excludes all protan subjects, except for those who pass the IH screening test.

**References:**

1. White TE, Rojas B, Mappes J, Rautiala P, Kemp DJ. Colour and luminance contrasts predict the human detection of natural stimuli in complex visual environments. *Biol Lett* 2017;13.

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

3. Bowmaker JK, Dartnall HJA. Visual pigments of rods and cones in a human retina. *Journal of Physiology* (London). 1980;298:501-511.

4. Cole BL. Colour blindness and driving. *Clin Exp Optom*. 2016;99:484-487.

5. 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.

6. 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.

7. 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

8. 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.

9. Barbur JL, Forsyth PM. *The effective contrast of coloured targets and its relation to visual search., Visual Search 2:* Taylor & Francis Ltd.; 1992:319-328.

10. D'Zmura M. Color in visual search. *Vision Res.* 1991;31:951-966.

11. Yamagishi N, Melara RD. Informational primacy of visual dimensions: specialized roles for luminance and chromaticity in figure-ground perception. *Percept Psychophys.* 2001;63:824-846.

12. Pinker S. Visual cognition: an introduction. *Cognition*. 1984;18:1-63.

13. Bowmaker JK. Trichromatic colour vision: why only three receptor channels? *Trends in Neuroscience.* 1983;6:41-43.

14. Dartnall HJA, Bowmaker JK, Mollon JD. Human visual pigments: microspectrophotometric results from the eyes of seven persons. *Proc R Soc Lond B.* 1983;220:115-130.

15. Paramei GV, Bimler DL, Cavonius CR. Effect of luminance on color perception of protanopes. *Vision Res.* 1998;38:3397-3401.

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

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

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

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

20. Verriest G. Further studies on acquired deficiency of color discrimination. *J Opt Soc Am.* 1963;53:185-195.

21. François J, Verriest G. On acquired deficiency of color vision with special reference to its detection and classification by means of the tests of Farnsworth. *Vision Res*. 1961;1:201-219.

22. Petzold A, Plant GT. Clinical disorders affecting mesopic vision. *Ophthalmic Physiol Opt.* 2006;26:326-341.

23. Pacheco-Cutillas M, Barbur JL, Edgar DF. Comparison of chromatic and achromatic loss of sensitivity in glaucoma. *Ophthalmic Physiol Opt.* 2002;22:574-574.

24. Pacheco-Cutillas M, Sahraie A, Edgar DF. Acquired colour vision defects in glaucoma - their detection and clinical significance. *Br J Ophthalmol.* 1999;83(12):1396-1402

25. 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.

26. Feitosa-Santana C, Paramei GV, Nishi M, Gualtieri M, Costa MF, Ventura DF. Color vision impairment in type 2 diabetes assessed by the D-15d test and the Cambridge Colour Test. *Ophthalmic Physiol Opt*. 2010;30:717-723.

27. Reffin JP, Astell S, Mollon JD. *Trials of a computer-controlled colour vision test that preserves the advantages of pseudoisochromatic plates.* Dordrecht, Netherlands: Kluwer Academic Publishers; 1991:69-76.

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

29. Rabin J, Gooch J, Ivan D. Rapid quantification of color vision: the cone contrast test. *Invest Ophthalmol Vis Sci.* 2011;52:816-820.

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

31. Birch J, McKeever LM. Survey of the accuracy of new pseudoisochromatic plates. *Ophthalmic Physiol Opt.* 1993;13:35-40.

32. Farnsworth D. *The Farnsworth dichotomous test for color blindness - Panel D-15*. New York: Psychological Corporation; 1947.

33. Fletcher R. Methods of examination of macular colour vision. A modified D-15 test. *Mod Probl Ophthalmol.* 1972;11:22-24.

34. Margrain TH, Birch J, Owen CG. Colour vision requirements of firefighters. *Occup Med (Lond).* 1996;46(2):114-24.

35. Birch J. *Diagnosis of defective colour vision*. 2nd Ed. ed. Oxford: Butterworth-Heinemann; 2001.

36. Wright WD, Holmes JG. A New Colour Perception Lantern. *Color Res Appl.* 1982;7:82-88.

37. Birch J. Performance of colour-deficient people on the Holmes-Wright lantern (type A): consistency of occupational colour vision standards in aviation. *Ophthalmic Physiol Opt.* 2008;28:253-258.

38. CIE. CIE. *Colours of signal lights*. Vienna Austria., Standard: S004/E-2001: Bureau Central de la CIE; 2001.

39. Birch J. Pass rates for the Farnsworth D15 colour vision test. *Ophthalmic Physiol Opt*. 2008;28:259-264.

40. Evans B, Rodriguez-Carmona M, Barbur JL. Assessment of chromatic sensitivity: 1. Visual signals that affect the outcome of color screening tests with emphasis on the Farnsworth D-15. *Color Res Appl.* 2020; submitted.

41. Birch J. Clinical use of the City University Test (2nd Edition). *Ophthalmic Physiol Opt.* 1997;17:466-472.

42. Franceschetti A. Die Bedeutung der Einstellungsbreite am Anomaloskop fur die Diagnose der eizelnen Typen der Farbensinnstorungen nebst Bemerkungen uber ihre Vererbungsmodus. *Schweiz Med Wochenschr*. 1928;52:1273-1278.

43. Jagle H, Pirzer M, Sharpe LT. The Nagel anomaloscope: its calibration and recommendations for diagnosis and research. *Graefes Arch Clin Exp Ophthalmol.* 2005;243:26-32.

44. Hurvich LM. *Color vision deficiencies.* In: Jameson D, Hurvich LM (eds), Handbook of Sensory Physiology. Berlin: Spriger-Verlag; 1972:582-624.

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

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

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

48. Chorley AC. A modified protocol for color vision screening using Ishihara *Aerosp Med Hum Perform.* 2015;86 747-751.

49. Webster MA, Miyahara E, Malkoc G, Raker VE. Variations in normal color vision. I. Cone-opponent axes. *J Opt Soc Am A Opt Image Sci Vis*. 2000;17:1535-1544.

50. Smith VC, Pokorny J. Chromatic-discrimination axes, CRT phosphor spectra, and individual variation in color vision. *J Opt Soc Am A.* 1995;12:27-35.

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

52. Cole BL. The handicap of abnormal colour vision. *Clin Exp Optom*. 2004;87:258-275.

53. Almustanyir A, Hovis J, Glaholt MG. Predicting the Farnsworth-Munsell D15 and Holmes-Wright-A lantern outcomes with computer-based color vision tests. *J Opt Soc Am A.* 2020;37:A1-A10.

54. Barbur JL, Rodriguez-Carmona M, Evans B. Assessment of chromatic sensitivity: 3. An efficient two-step color assessment protocol. *Color Res Appl.* 2020; submitted.

55. Barbur JL, Rodriguez-Carmona M, Evans S, Milburn N. *Minimum Colour Vision Requirements for Professional Flight Crew. Recommendations for new colour vision standards.* Paper 04/2009; 2009.

56. 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.

57. Barbur JL, Rodriguez-Carmona M, Morgan MJ. "Double-Blindsight" in Human Vision. *Investig Ophthalmol Vis Sci.* 2002;43:3909.

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

**Author Biographies:**

**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.

**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 Colour Assessment and Diagnosis test and more recently, the Colour Vision Screener test. These and other

Advanced Vision and Optometric Tests were developed over several years and found important applications in both clinical work and in visually-demanding occupations.