Assessment of Current and Next Generation of Colour Vision Tests for Occupational Use

by

Ali Almustanyir

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Vision Science

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner   NAME: Michael Crognale

                                           Title: Professor

Supervisor(s)   NAME: Jeffery Hovis

                                           Title: Associate Professor

Internal Member   NAME: Vengu Lakshminarayanan

                                           Title: Professor

                                           NAME: Daphne McCulloch

                                           Title: Professor

Internal-external Member   NAME: Ian McKenzie

                                           Title: Adjunct Associate Professor & Associate

                                                  Chair Undergraduate Studies, Aviation &

                                                  Geomatics

AUTHOR'S DECLARATION

This thesis consists of material all of which I authored or co-authored: see Statement of

Contributions included in the thesis. This is a true copy of the thesis, including any required

final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The description and review of colour vision tests on Chapter 1, method Chapter, and Chapter 8 appeared in a technical report by Hovis and Almustanyir (2017). I wrote the first drafts of all Chapters and I did the necessary edits.

Abstract

Introduction: Individuals with congenital colour vision deficiencies are at a greater risk of making an error in colour judgment. Computerized colour vision tests are now available to screen for colour vision defects and quantify the severity of their defect.

Purposes: The first experiment compared chromatic thresholds measured on different computerized colour vision tests for colour-normals (CVN) and colour vision defective (CVD) using a common scale. The next experiment evaluated whether dichromatic transformations of the Farnsworth D15 (F-D15) and the ColorDx D15 colors could predict the actual arrangements. The third study evaluated the within-session and between visits repeatability of each of the D15 tests. The last studies determined which one of the newer computer-based tests would replace the Holmes-Wright Type A Lantern (HWA), F-D15 and the CN Lantern. Lastly, the results of a short survey of how a colour vision deficiency affected their lives are presented.

Methods: Sixty CVN subjects and 68 CVDs were tested with the Oculus HMC anomaloscope, Psuedoisochromatic tests (the Ishihara, Hardy, Rand, Rittler (4th ed), ColorDx PIP), computrized tests (CAD, Rabin Cone Contrast Test (RCCT), Cambridge Colour Vision Test (CCT), Landolt C Cone Contrast (LandC)), Lantern tests (HWA, CN Lantern), and arrangment tests (F-D15, ColorDx D15).

Results: Discrimination ellipses measured in the CIE colour space with the CAD were significantly larger than the ellipse areas measured by CCT for CVNs, protans and deutans.

For the tests that measure vector length, there were significant interactions between the three tests and the different subject groups. In general, the dichromatic transformation of the D15 tests provided reasonable predictions of the actual dichromatic arrangements. Both D15 tests studied showed that each test was highly repeatable within and between sessions. As expected, the ColorDx D15 had the highest level of agreement with the F-15, although it was not as challenging as the F-15. The LandC and CAD had significantly higher levels of agreement with the HWA compared with the other tests. Nearly all the CVD failed the CN Lantern test at 4.6m, but the number of errors decreased and the pass rate increased as the viewing distances of the CN lantern was shortened. At the 2.3 m distance, the agreement values for the HRR, ColorDx, and CAD were similar and considered as good. These agreement values decreased as the viewing distance decreased further. The results of the survey study showed that the percentage of dichromats who reported difficulties with colours was more than double of the percentage of anomalous trichromats, but lower than reported by other surveys.

Conclusions: The difference between the CAD and CCT colour discrimination ellipses was likely a result of the different number and spacing of the chromatic vectors that were sampled in each program. The differences between the three tests that measured the vector length could be due to the luminance masking noise, monitor artifacts, different angular sizes, and psychophysiological procedures. The RCCT and LandC tests did not show a potential ceiling effect in estimating CVD thresholds, however, the LandC was preferred over the RCCT

because it can measure chromatic threshold for CVN. The dichromatic predictions of the D15 tests suggest that this model may be useful in predicting the performance of dichromats on other colour-related tasks. The decrease in errors on the CN Lantern was likely due to the increase in brightness of the test lights with decreasing test distance. Although everyone with a colour vision defect will likely to fail the CN Lantern at 4.6 m, individuals who fail either version of the D15 will almost certainly fail the CN Lantern at 4.6 and 2.3m. A mild classification on the HRR and ColorDx PIP provides a reasonable prediction of who will pass the CN lantern at the shorter distances.

Acknowledgements

I would like to thank my supervisor, Dr. Jeffery Hovis who was more than helpful with his expertise and time. I would like to thanks my committee members Dr. Vasudevan Lakshminarayanan and Dr. Daphne McCulloch for their assistance.

Special thanks to the co-op students Zach Barnes, Jonathan Siu, and Danny Fusco for their help.

I would like to thanks Varadharajan Jayakumar for helping in photographing the tests.

I would like to acknowledge and thanks the School of Optometry and Vision Science, University of Waterloo for allowing me to conduct my research.

Grateful thanks to my fellow graduate students, faculty members, and staff who supported me during my time at the School.

I would like to thanks the computer technicians, the Optometry Library staff, and the graduate coordinator (Stephanie Forsyth) for their help and support.

I would like to acknowledge the Defence Research and Development Canada and the Canadian Institute for Military and Veteran Health Research for their financial support and the United States Air Force OBVA Laboratory for loaning us the LandC test.

A special thanks to my sponsor, King Saud University and the Saudi Arabia Cultural Bureau in Ottawa for their financial support.

Dedication

I would like to dedicate my thesis to my parents, my brothers and sisters. To my wife and sons (Mansour and Rakan). I could not do this without your support and encouragement to finish this mission from the first day to the moment of my defense. I would also like to dedicate this thesis to all of my friends for their continued support.

Table of Contents

xiii

List of Figures

xx

List of Tables

Chapter 1
Introduction

1.1. Normal colour vision

1.1.1 Anatomy Overview

The retina of the eye is responsible for transforming radiant energy from the visible spectrum into neural impulses and for early neural processing of visual sensation. This process starts with the absorption of the photons by the photopigment located in the outer segments of the photoreceptors (Oyster, 1999). Figure 1 represents a simple anatomy of the retinal region. The photoreceptors synapse to the bipolar cells, which send the signals into the retinal ganglion cells. Horizontal cells mediate lateral neural interactions in the outer retinal layers through their connections with the cones and bipolar cells; whereas amacrine cells mediated lateral interactions in the inner retinal layers via their connections with the ganglion and bipolar cells (Schwartz, 2009). The neural signals leave the eye via the optic nerve to the optic chiasma. The pathway then becomes the optic tract and goes to the lateral geniculate nucleus (LGN). The ganglion cell consists of three major classes of neurons, which are the midget ganglion cells, parasol ganglion cells, and bistratified ganglion cells. The LGN is divided into three distinct regions. The two most ventral layers are magnocellular layers, which contain the large neurons called magnocellular (magno or M cell) cells. The dorsal four layers are the parvocellular layers, which have smaller neurons called parvocellular (parvo or P cell) cells. The area between these layers contains neurons called koniocelluar (konio) cells (Schwartz, 2009).

1

The three classes of ganglion cells synapse in different layers of the LGN. The midget ganglion cells synapse the P cells to form the parvo retinogeniculate pathway whereas the parasol ganglion cells synapse the M cells to form the magno retinogeniculate pathway. The bistratified ganglion cells projection synapse the konio cells to make the konio retinogeniculate pathway (Schwartz, 2009). The parvo pathway encodes fine spatial detail and red-green chromatic information. The the magno pathway codes information regarding low contrast, large objects and motion. The konio pathway plays a major role coding the blue-yellow chromatic information (Schwartz, 2009).
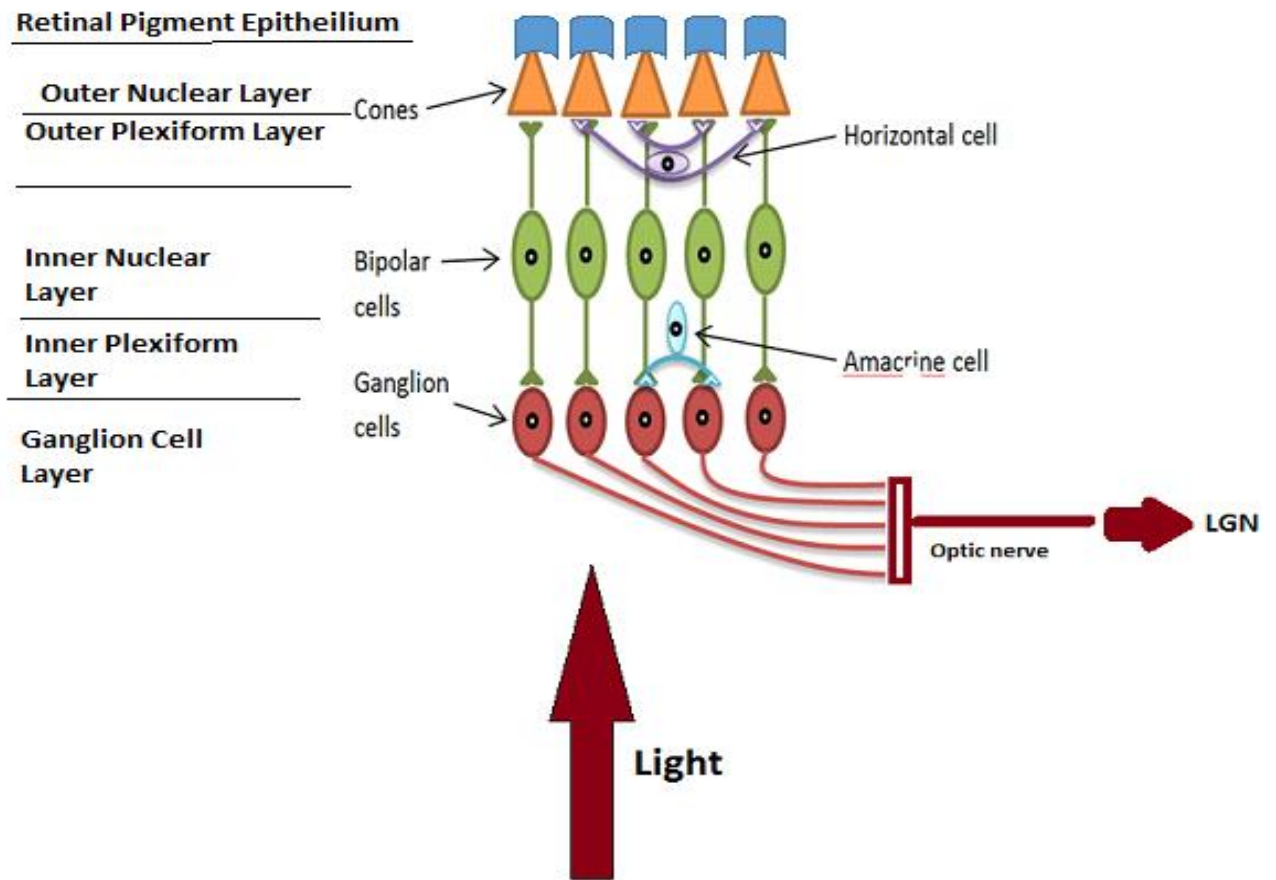
Figure 1. Simplified schematic of the cellular organization in the retina. The light travels from the inner layers of the retina to the outer layers until it reaches the photoreceptors. The photoreceptors absorb quanta and convert the energy into electrical signals. The photoreceptors synapse with the bipolar cells. The bipolars stimulate the ganglion cells that send the signals to the LGN through the optic nerve.

The photoreceptors consist of two groups: rods and cones. Rods are the receptors that mediate night vision and do not have a primary role in coding colour. The cones mediate daytime vision, which includes the ability to see fine details and colours (Schwartz, 2009). There are three classes of cones. There are long wavelength sensitive cones (L-cone), Medium wavelength sensitive cones (M-cone), and Short wavelength sensitive cone (S-cone). Although the cones are most numerous in the foveal area, the distribution of the three types of cones is random (Mollon and Bowmaker, 1992). The mean L: M: S cone densities are in the ratio of approximately 40:20:1 (Kremers et al., 2000).

1.1.2 Ocular Media and abnormality in colour vision

When the light enters the eye, a proportion of it is lost due to reflection, scattering, and absorption by the ocular media (cornea, aqueous humour, crystalline lens and vitreous humour). The amount of absorption and scattering are wavelength dependent (Fletcher & Voke, 1985). Figure 2 shows the total transmission of the ocular media as a function of wavelength. Most of the absorption is at the blue and violet wavelengths (Fletcher and Janet, 1985).
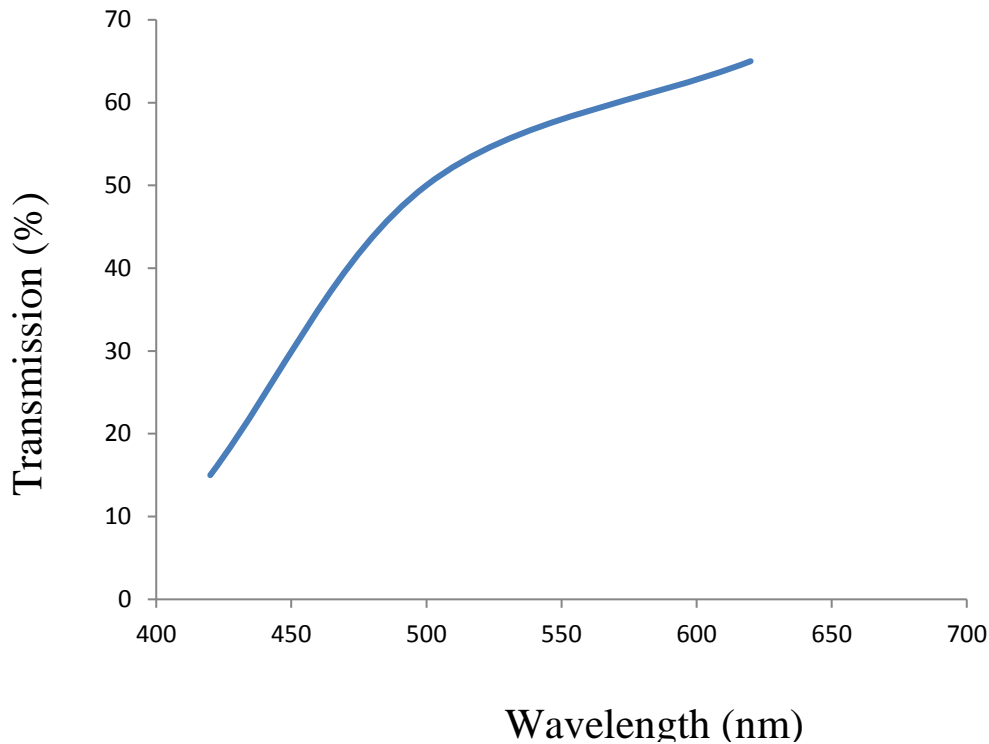
Figure 2. Transmittance of the ocular media (cornea, aqueous humour, crystalline lens and vitreous humour (Fletcher and Voke, 1985).

In the foveal area where the cones are densest, there is a yellow coloured pigment distributed throughout the retinal layers referred to as the macula pigment (MP) (Remington, 2011). The (MP) could protect the retina from harmful blue light wavelengths and/or reduce blur due to chromatic aberration (Fletcher and Voke, 1985). MP absorbs short wavelength light in the region of 400 to 540-nm, with maximum absorption near 460 nm. The MP consists of three carotenoids: (lutein, zeaxanthin, and meso-zeaxanthin) (Remington, 2011).

Figure 3 shows the human macular pigment absorption. Light absorption by the ocular media

will affect the light perception, especially in at the shorter wavelengths.
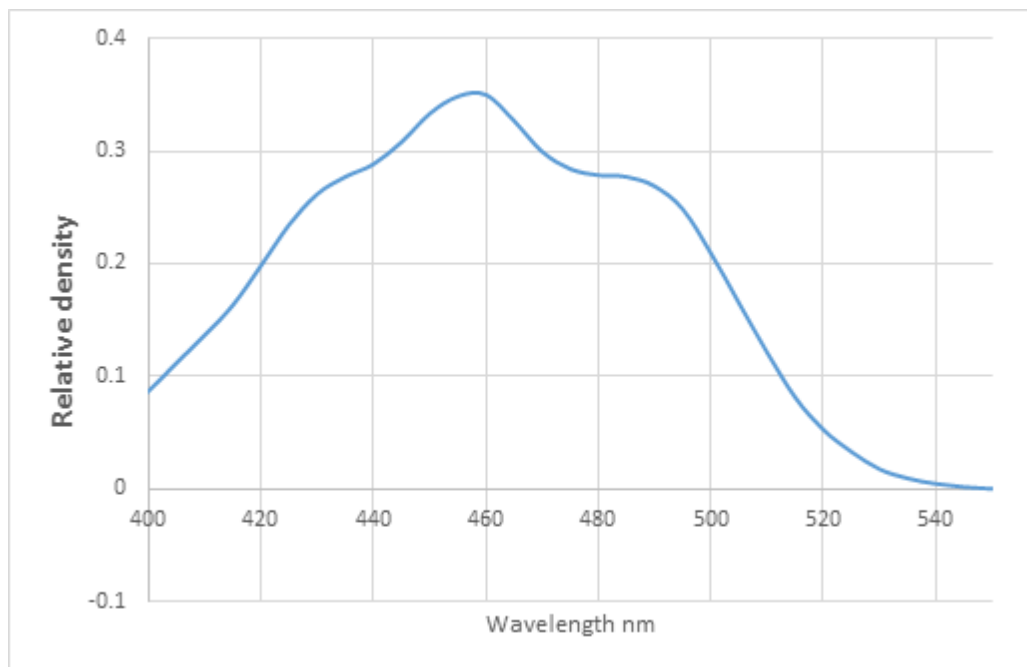
**Figure 3 Macular pigment densities with a different wavelength of the spectrum (Fletcher &
Voke, 1985).**

Colour discrimination improves from birth to about the early 20s and then begins to

degrade slowly after 30 years of age (Fu et al., (2009); Fiorentini et al., (1996)). Generally,

older individuals show a loss in blue-blue-green discrimination (Fletcher and Voke 1985).

Although diseases of the visual system increase with age and could be responsible for this

colour discrimination loss, the main factors are pupil miosis and the yellowing of the

crystalline lens (Fletcher and Voke, 1985). Some clinical colour vision tests can measure these age-related in colour discrimination.

1.1.3 Colour

Colours are usually described by three attributes: hue, saturation, and brightness. These three variables are wavelength dependent. Figure 4 shows a schematic diagram of the visible spectrum. This visible spectrum extends from 380nm to 780nm. The approximate   hues of the visible from the short to the long wavelengths are violet, blue, green, yellow, orange, and red (Schwartz, 2009).

Violet	Blue	Green	Yellow	Red

380   430   480   530   580   630   680   730   780

Wavelength
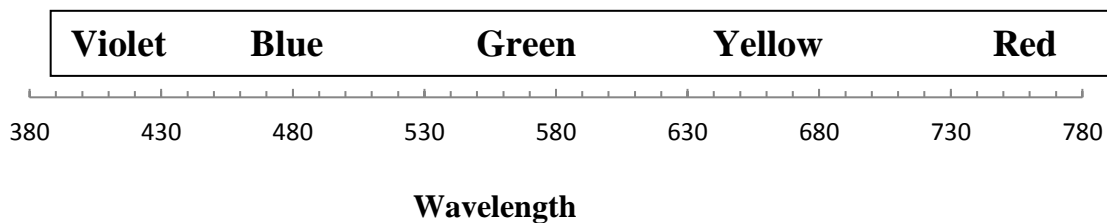
Figure 4. Schematic diagram of the visible spectrum from 400nm (violet colour) to 700 nm (red).

1.1.3.1 Hue

Hue perception arises from the combinations of four hues: blue, green, yellow and red. These four colours are called unique colours. These four colours are organized in two pairs: red-green and blue-yellow. Hues do not look reddish-green or yellowish-blue. The red and green are in opposition as are blue and yellow (Pokorny et al., 1979). Wavelength discrimination

can be described as the amount of wavelength change that is required to notice a difference in hue when the lights are at equal luminance.

1.1.3.2 Saturation

Saturation can be described as the vividness of the colour. The saturation scale varies from very saturated colour (very little white) to completely desaturated colour (i.e. white). For example, adding white light to spectral light will make the colour desaturated or appear more pastel. For spectral lights at equal luminance, saturation is higher at short and long wavelengths relative to wavelengths in the yellow-green wavelengths (Pokorny et al., 1979). Saturation discrimination is often determined by is the minimum luminance of spectral light that is necessary to make white appears just noticeably different while keeping the luminance of the mixture constant (Pokorny et al., 1979). The ratio of luminance of the spectral colour (L_λ) divided by the luminance of the mixture of the spectral colour and the luminance of white (L_w) is referred to as the colorimetric purity, and it is reciprocally related to the saturation of L_λ. If the purity is low, then very little of the spectral light is added to change the appearance of the white reference, which indicates that the spectral light had a high level of saturation and saturation discrimination was very good for this test light.

1.1.3.3 Brightness

Brightness is the perception of the light intensity. The measurement of brightness is challenging because it is a perceptual phenomenon and brightness perception may be influenced by the viewing condition, state of adaptation, ambient light, and the immediate

surround of target (Pokorny et al., 1979). Nevertheless, brightness often quantified by measuring the luminance of an extended source. Luminance is the light intensity per unit of a projected area of the light source or object. The units are candela per meter square (cd/m²). Generally, the terms brightness and luminance are interchangeable. However, they are not exactly the same. For example, brightness is a power function of luminance with an exponent less than 1.0 so that if the luminance of a light is increased for example by 5 times, an individual might report that the increase in brightness is less than 5 times. Colour discriminate is best at the photopic levels (above 3.0 cd/m²). When the light level is reduced to a mesopic level (between 0.001 and 3.0 cd/m²), there will be some loss in colour discrimination. At the lower scotopic levels (less than 0.001 cd/m²), colour discrimination reduced further, and the visual perception may mediated by primarily rod input. (Pokorny et al., 1979).

1.1.4 Basic Color Perception Models

Many colour vision models have been proposed to explain colour appearance; however, the more recent ones are based on the Hering-Hurvich-Jameson colour vision model (Jameson and Hurvich, 1955; Hurvich and Jameson, 1955). Figure 5 shows a schematic of the model for normal colour vision. The first stage of the process begins at the cones level (circles in Figure 5). The L-cone, M-cone and S-cone provide input into 3 post-receptor channels. Two channels code chromatic information and one channel codes achromatic information. One chromatic channel, the blue-yellow channel, signals that the stimulus is either blue or yellow.

Its output is calculated by subtracting the sum of the L cone and M-cone responses from the S-cone. A positive output from this channel signals blue and negative output signals yellow. If there is no output, the colour appears neither blue nor yellow. The second chromatic channel, the red-green channel, signals that the stimulus is either red or green by subtracting the M-cone response from the sum of the L and S cone responses. Positive values signal red and negative values signal green. When the output is zero, the colour appears neither red nor green. The third channel is the achromatic channel. It codes either whiteness or blackness. The achromatic channel response is determined by the sum of the L-cone and the M-cone. The achromatic channel is also referred to as the luminance channel. If there is no response from the chromatic channel, the light appears either white or black. Positive output from the achromatic channel signals white and negative output signals black. (Jameson and Hurvich, (1955); Hurvich and Jameson, (1955)).

The output of the blue-yellow channel shows that there is a large blue response at short wavelengths, which goes to zero at 500 nm and becomes a yellow response at longer wavelengths. The red-green channel shows large red responses at the shorter wavelengths and reaching zero near 475 nm, green response from 475 nm to 575 nm and becomes a red response at the longer wavelengths. Because there is no output of the red-green channel at a wavelength near 475 nm and the blue-yellow channel's response is blue, this light appears 100% blue (neither red nor green). Moreover, a wavelength near 575 nm appears 100% yellow and a wavelength near 500 nm appears 100% green. Because there is both a red signal from the red-green channel and a small yellow signal from the blue-yellow channel, longer

10

wavelengths appear reddish orange. A light that appear 100% red produced by mixing a short

wavelength (<475) with a long wavelength. .
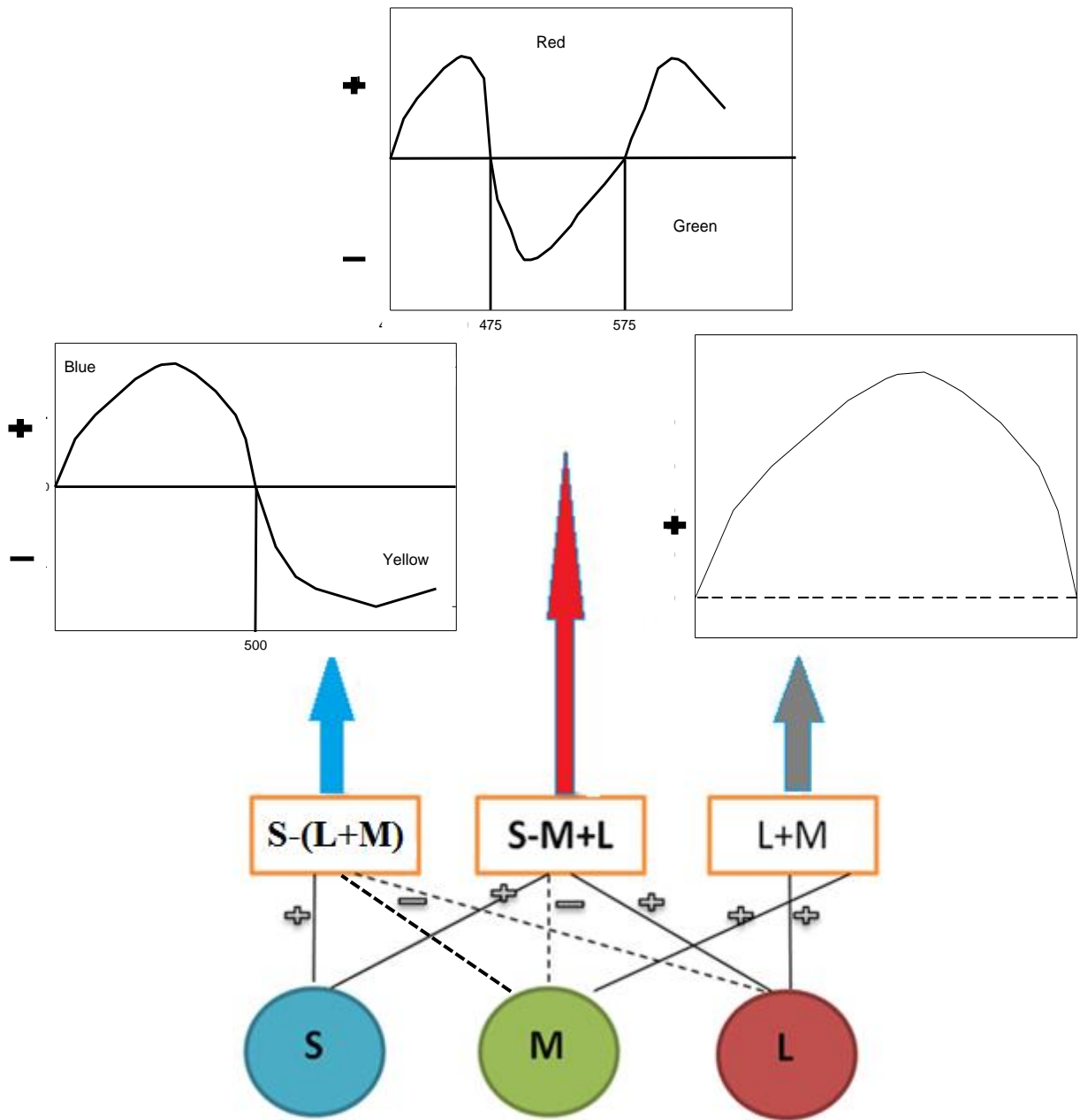
Figure 5. The cones give inputs (excitatory or inhibitory) into two chromatic channels and one achromatic channel. Note that in the above figure that the inhibitory inputs are represented by dash line (----) and the excitatory inputs by continue line (___). The addition and subtraction of the receptor inputs will determine the output of the post-receptor channels, which can be used to predict the hue of a stimulus.

1.1.5 CIE colour system

As mentioned previously, the first stage in translating optical radiation into colour perception is the absorption photons by the photopigment in the cone outer segments. However, once a photon is absorbed, the response of the cone is independent of wavelength. That is, the number of photons absorbed, and not the wavelength, of the light determines the cone responses. This is the principle of univariance. As a result, two lights that differ in spectral composition (i.e. different wavelengths and energy) may look identical because the cone responses to the two lights are identical. This is called metameric match because the two lights look identical but do not have the same spectral composition (Schwartz, 2009).

Figure 6 illustrates the idea of the metameric match. The test wavelength (t_λ) fills the top part of the bipartite filed. The red (r_λ), green (g_λ), and blue (b_λ) primaries are on the bottom portion. The amount of the three primaries (r_λ, g_λ, b_λ) is adjusted to make a perfect match with the test light (Schwartz, 2009). Technically, one of the primaries may have to be used as a desaturant and combined with the test light in order to make an identical match between the two halves of the stimulus field. The primaries usually located in the short, medium and long wavelength regions of the spectrum and so they are often called red, green and blue. The

12

result that all colours can be matched by the appropriate amount of three primaries has led to a colour specification system based on colour matches.
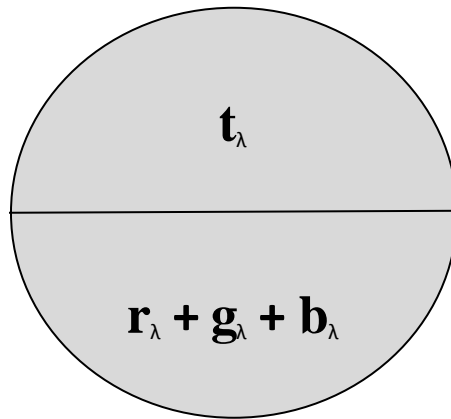
Figure 6. Colour match based on the R, G, B system. A test light (t_λ) is given by the relative amount of three primaries (r_λ, g_λ, b_λ) needed to make a colour match.

Figure 7 shows the relative amounts of the 3 primaries required to match each wavelength listed on the x-axis. These results are from the International Commission on Illumination (CIE) standard observer system (Schwartz, 2009). The amount of each of this primary that is needed to make a colour match is the tristimulus value. Negative tristimulus values indicate that the primary was used as desaturant. For example, the negative r_λ values

for 450 to 550 nm indicate the amount of the r_λ that had to be added to the test light in order to obtain a match with the other two primaries.
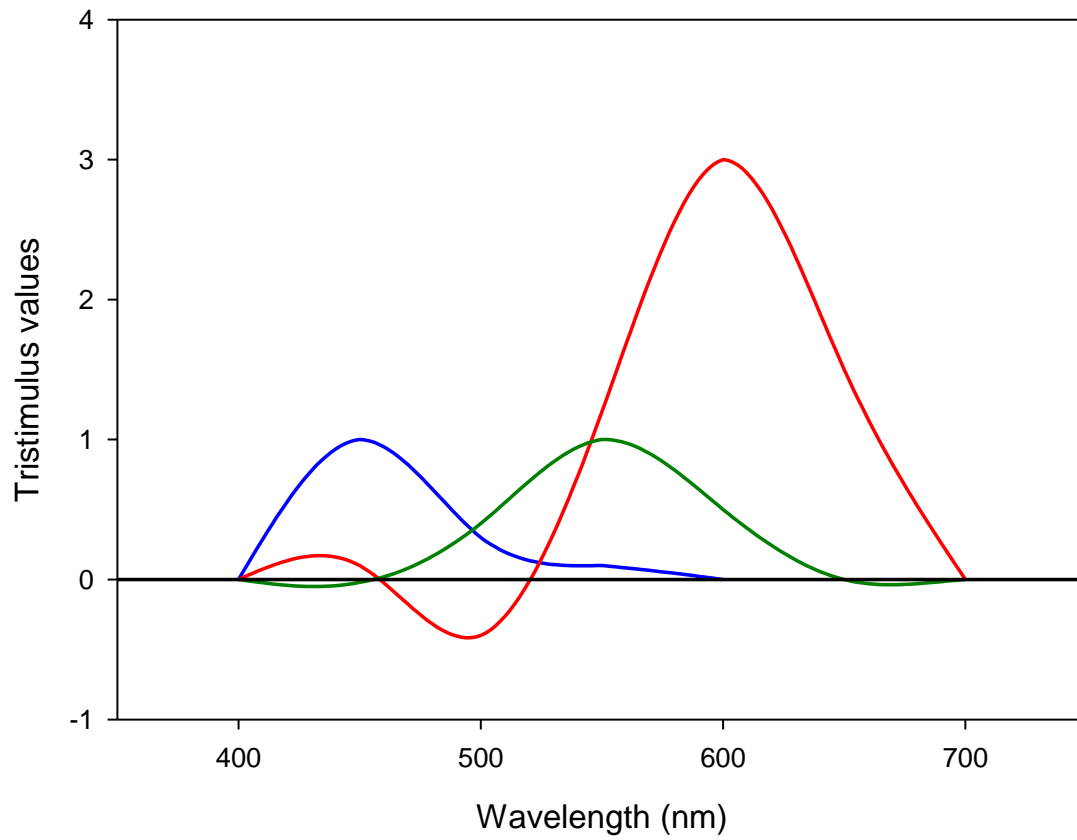
Figure 7. Colour matching function based on the R, G and B system. The red, green and blue curves represent the R, G, B primaries respectively.
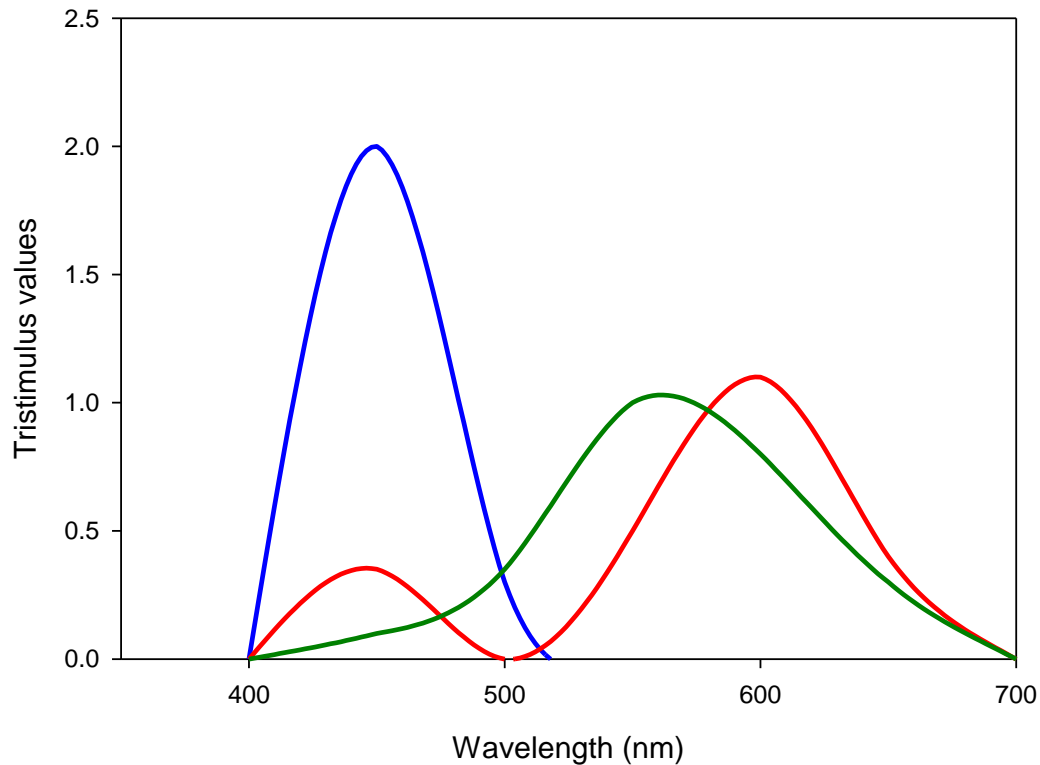
Figure 8. Colour matching function of the 1931 CIE diagram for the three primaries X (red), Y (green), Z (blue). These primaries were obtained from a transformation of the R, G and B primaries.

The RGB colour matching functions were transformed in order to eliminate negative numbers and have one function identical to the human spectral brightness sensitivity function. Figure 8 shows the results. The three primaries are called X, Y, and Z because they are not actual lights, but mathematical transformations of R, G, and B primaries. Because the relative amount of the 3 primaries should not change with brightness as long as there is no photopigment bleaching, colour specification can be further simplified by converting the

tristimulus values to the chromaticity coordinates. The chromaticity coordinates are designated by lowercase x, y and z and are calculated from the tristimulus values for a given light by the following equations:

$$x = \frac{X}{X + Y + Z}$$

$$y = \frac{Y}{X + Y + Z}$$

$$z = \frac{Z}{X + Y + Z}$$

where $x + z + y = 1$

This conversion allows the results to be plotted in a two-dimensional system. Figure 9 shows the 1931 CIE chromaticity diagram. It shows the relative amounts primaries that are required to make a colour match. The curved line is called the spectral locus, and it is where the spectral lights (i.e. monochromatic lights) are located. All lights fall either on or within, the spectral locus. The straight line between 400 nm and 700 nm is referred to as the line of purples. Colours along this line represent mixtures of 400 nm (blue) and 700 nm (red). Technically, the CIE diagram does not specify appearance. It only specifies the relative amounts of the 3 primaries required to obtain a match. Nevertheless, if one assumes that the observer is adapted to a neutral light and the light levels are representative of office lighting, then one can add colour names to the diagram to indicate the approximate locations of the different colours within the diagram.

A fundamental attribute of the CIE chromaticity diagram is that if two colours are mixed together, the resulting colour will lie on the line joining the two original colours. In Figure 9, the colour created by mixing A with B will fall on the straight line between the two points and its location along the line will be determined by the relative amounts of A and B. If more B is in the mixture, then the resulting colour will fall on the line closer to B.
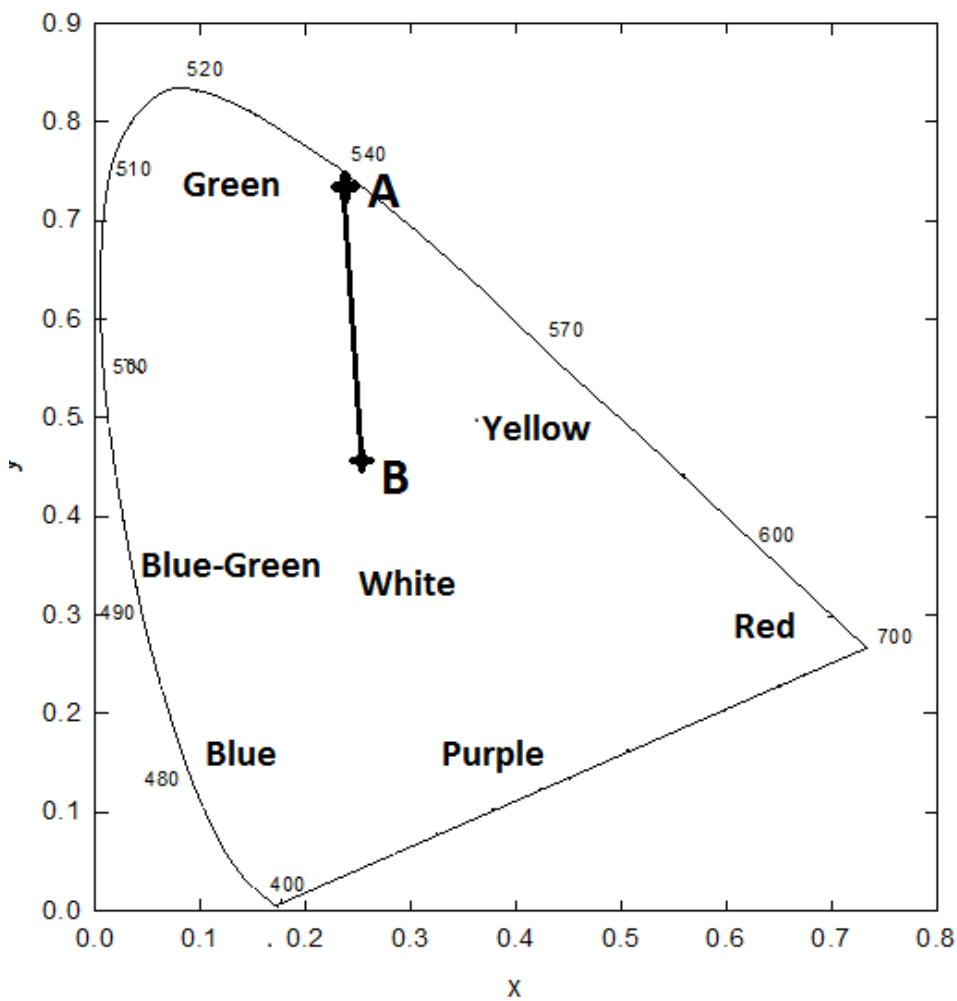
Figure 9. CIE 1931 (x,y) chromaticity diagram. x, y, and z are equal to each other at the equal energy white. The numbers along the curve are spectral wavelengths in nm. All the feasible colours are within this diagram. The line between A and B is the colour matching between the two points. The colour A is close to the spectrum locus, and point B is close to the centre of the diagram.

1.1.5.1 Excitation purity and Dominant wavelength

As mention previously in section 1.3, colour appearance is specified by the three basic attributes of colour are hue, saturation, and brightness. Because the CIE diagram specifies the relative amounts of the primaries, brightness information is lost. Although hue and saturation cannot be specified on the CIE diagram, they can be approximated by using dominant wavelength and excitation purity (Pokorny et al., 1979).

In Figure 10, the point C on the CIE chromaticity diagram is the white light source illuminating a coloured object. The chromaticity coordinates of the colour are at point F. Colour F can also be obtained by the additive mixture between the wavelengths at point D (487 nm) with C. Point D is called the dominant wavelength of the colour F because it is the spectral light that is mixed with C in order to match colour F. Because the wavelength is 487 nm, one can infer that the colour of F is bluish. Obtaining the dominant wavelength at point E involves a few more steps because the line extending from the white reference through E intersects the line of purples at B so that there is no corresponding wavelength. In this case, the line will be extended in the opposite direction to get the spectral locus at point A (510 nm). The corresponding wavelength at point A is actually the complementary wavelength to lights E and B because when A is mixed with either E or B in the appropriate amount, the

18

mixture will match the reference white. In the case of any purple colour, the dominant

wavelength is actually the complementary wavelength.
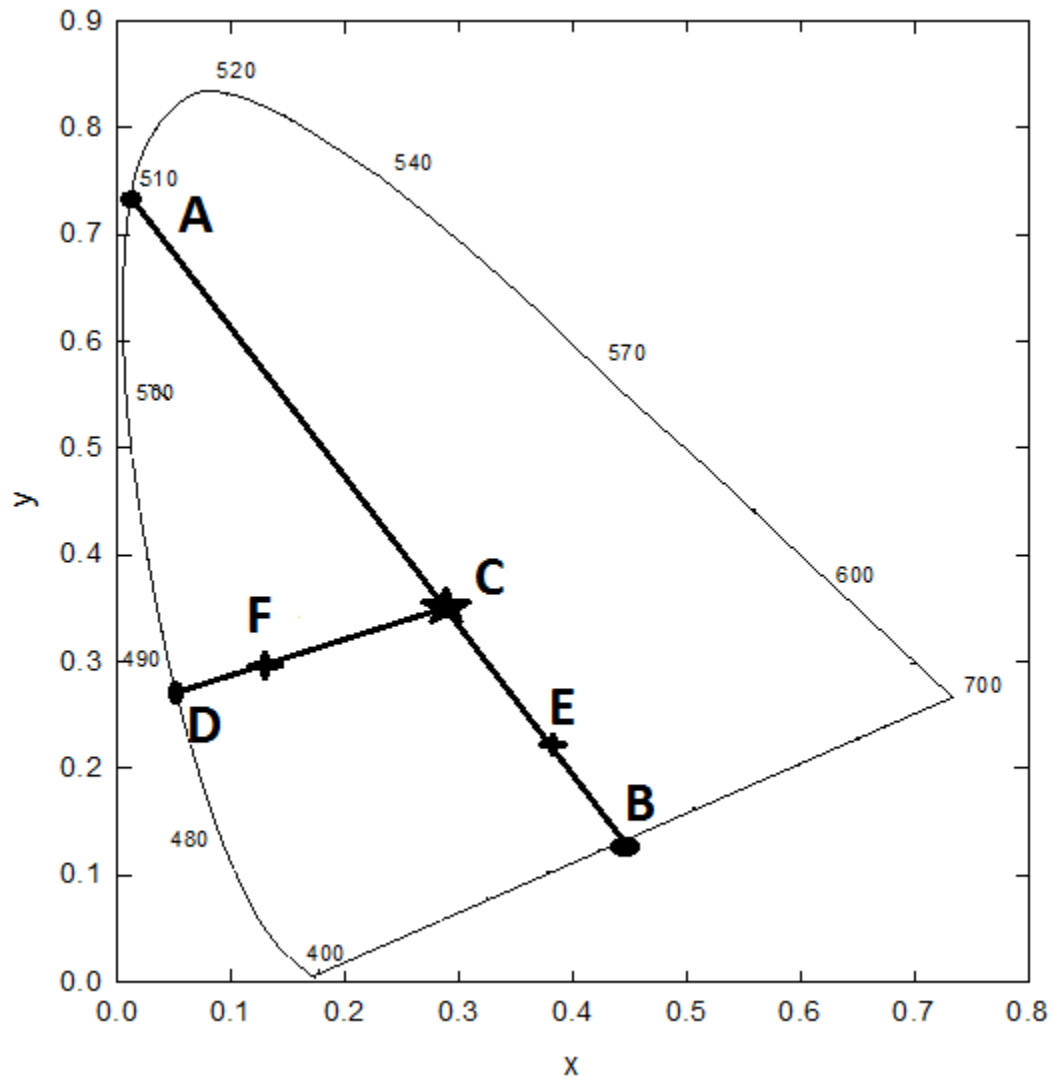
Figure 10. The CIE diagram with the dominant wavelength for colours on the spectral locus and the complementary colour on the CIE diagram. The point C represents daylight. The point F can be obtained from the mixture of the daylight and 487 nm (point D). Point E can be obtained by the mixing point B with its complementary colour point A (approximately 510 nm).

In the CIE diagram, excitation purity (Pe) is defined as the ratio of the distance from a specific white stimulus point to a given point, divided by the distance from the white to the dominant wavelength (or point on the line of purples) (Pokorny et al., 1979). In figure 11 the excitation purity is the ratio of the distance from C to R (labeled as a) to the distance from C to the dominant wavelength (labelled as a+b). It can be calculated as

$$Pe = \frac{a}{a + b}$$

When the value of Pe is nearly 1.0 then the colour is considered as having its maximum saturation and lies on the spectral locus. However, when the value of Pe is near zero, then the colour tend to be desaturated, and the colour point is near the white stimulus.

The diagram includes the formula:

$$Pe = \frac{a}{a+b}$$
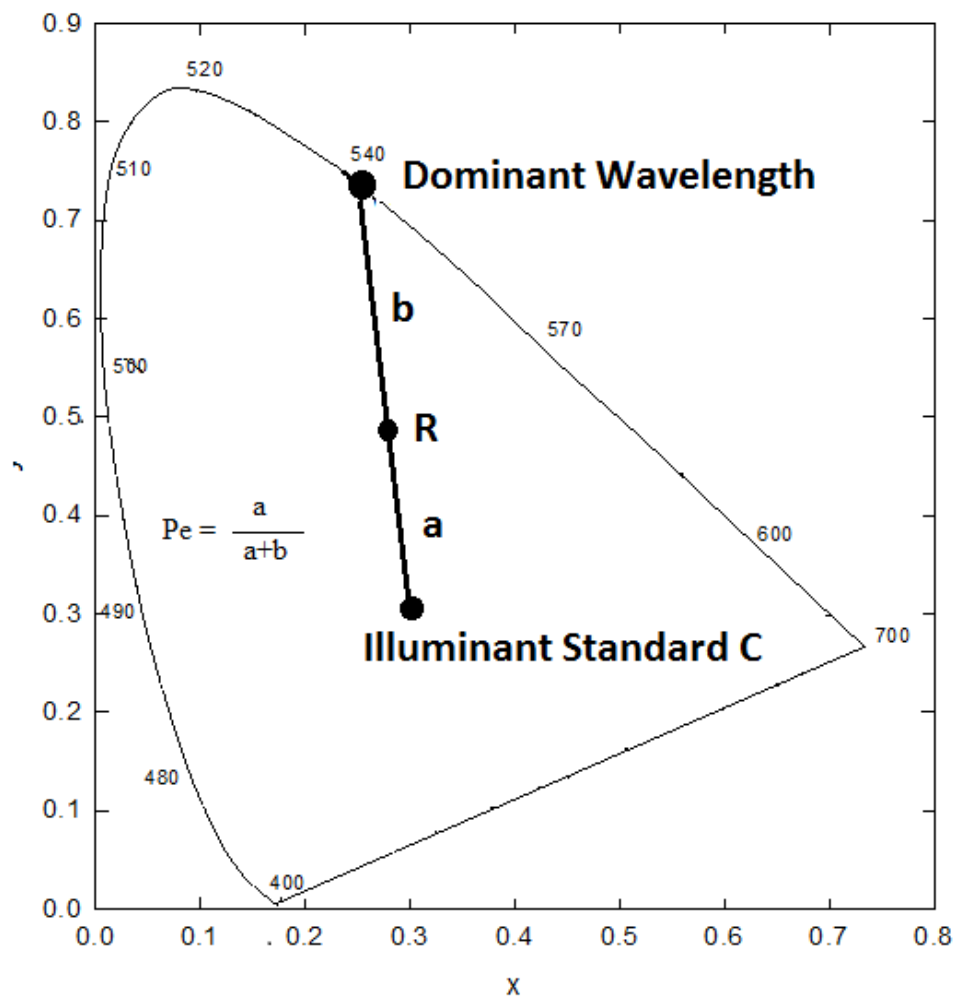
Figure 11. The excitation purity relative to the Illuminant C on the CIE diagram.

1.2 Colour Vision deficiency

Colour vision deficiency is described as having difficulty distinguishing colours, having

colour matches that are outside the normal range, or both. The severity of the deficiency can

range from mild to severe depending on the nature of the underlying problem. Colour vision

deficiency can be divided into congenital and acquired colour vision deficiencies (Shin et al.,

2007). The congenital (X-linked defeciency) defect is inherited, and it occurs independently

of any other problem with a person's visual system. The colour vision loss remains relatively

stable throughout life (Shinomori et al., 2016). Acquired colour vision deficiencies are

manifest with an underlying disease or disorder. The disease or disorder could be inheritable,

but the key difference between congenital and acquired is that some other aspect of visual

function is associated with the loss of colour discrimination. An acquired defect can progress

depending on the underlining condition (Schwartz, 2009; Pokorny et al., 1979). Acquired

defects are less prevalent in the young adult population than the congenital colour vision

defects (Schneck et al., 2014).

1.2.1 Congenital colour vision deficiency

Congenital colour vision deficiencies are classified as either red-green or blue-yellow based

on the colours that are likely to be confused. Red-green defects are X-linked recessive and

relatively common, affecting approximately 8% of Caucasian males and 0.5% of Caucasian

females. The prevalence is lower in other ethnic groups (Delpero et al., 2005; Birch, 2012).

Congenital blue-yellow defects are autosomal dominant with incomplete penetrance (Deeb,

2004). These defects are very rare, with a prevalence of approximately 0.1% in the general

population (Pokorny et al., 1979; Delpero et al., 2005). Individuals with red-green defect

confuse colours along the L and M cone confusion axis (red, orange, yellow and green),

whereas individuals with blue-yellow defect confuse colours along the S and L+M cone

confusion axis (violet, grey and yellow-green). Within these two broad categories, the defect can be divided into dichromat and anomalous trichromatic based on the number of primaries required to match a coloured light. Dichromat individuals require only two primaries to make a match and anomalous trichromatic require three primaries, but the amounts of the primaries are different from the colour vision normal (CVN) population. On average, dichromats also have worse colour discrimination than anomalous trichromats (Pokorny et al., 1979).

Red-green colour defectives (CVD) can be further classified based on whether the M-cone or L-cone photopigment is missing, or different, from the CVN population. Deuteranomalous individuals have an altered M-cone photopigment that has a peak absorption closer to the L-cone and deuteranopes are missing the M-cone photopigment. Protanomalous individuals have an altered L-cone photopigment which has a peak absorption closer the M-cone and protanopes are missing the L-cone photopigment (Pokorny et al., 1979).

The dichromat blue-yellow defect is referred to as tritanopia, and the blue-yellow anomalous trichromat is referred to as tritanomaly. In both cases, there is a problem with S-cones. In tritanopia, the S-cone pigment is non-functioning, whereas in tritanomaly there is a partial functioning of the S-cone (Deeb, 2004).

Because most clinics do not have an anomaloscope to perform colour matches, they cannot distinguish between anomalous trichromats and dichromats. For that reason, deuteranomalies and deuteranopes are classified clinically as deutans; protanopes and protanomalies are classified as protans and tritanopes and tritanomalies are classified as

tritans. Many colours confused by protan individuals are similar to the colour confused by deutan individuals (Pokorny et al. 1979). One of the major differences between these two subtypes is that protans have a decreased brightness sensitivity to red lights and deutans have a normal brightness sensitivity to red light.

1.2.2 Acquired colour vision deficiency

Acquired colour vision deficiencies are more common in the elderly (Schneck et al., 2014). This is due to the fact that the prevalence of visual disorders also increases with age. Glaucoma, cataract, macular degeneration, and diabetes are the most common causes of acquired colour vision defects (Pokorny et al., 1979). Acquired colour vision defects can be divided into three types. The most common type is Type III blue –yellow defect. Observers with this type of defect have discrimination losses similar to a congenital tritan defect This defect is associated with macular degeneration, glaucoma, diabetes, nuclear cataract, and some optic nerve disorders (Pokorny et al., 1979). There are two types of red-green acquired colour vision defects. Type I red-green defect is a result of a photoreceptor/retinal pigment epithelium disease or disorder. This type of defect is defined as a loss in colour discrimination along the red-green axis. Individuals with this type of defect tend to require more red in a mixture of red plus green to match a yellow reference. Type II acquired red-green is associated with optic nerve diseases such as optic atrophies and optic neuritis (Pokorny et al., 1979). Patients with this type of defect tend to require more green in a

mixture of red plus green to match a yellow reference, along with reduced hue discrimination along the red-green axis.

1.2.3 Colour confusion lines and congenital colour vision deficiency

A common method for determining whether colours are confused by individuals with a congenital defect is to add the dichromat lines of confusion to the CIE diagram. A single line represents all colours that require the same ratio of the two dichromat primaries in order to make a colour match. As a result, if all the colours along the line have the same brightness, then they will appear identical to each other to a dichromat. The straight lines in Figure 12, 13 and 14 are the colour confusion lines for the protanope, deuteranope and tritanope respectively (Vos and Walraven, 1971; Walraven, 1974; Pitt, 1944; Wright, 1952). For example, in Figure 12, the purple and green colours plotted in the diagram will appear identical to a deuteranope if they are equal in brightness because they fall among the same confusion line. The line of confusion that passes through a reference white is referred to as the neutral axis because all the colours along this line will appear identical to the reference white. The spacing between the lines of confusion is determined by hue discrimination. That is, each line represents a just noticeable different colour for the respective dichromat. The discrimination of the anomalous trichromats is qualitatively similar to their respective dichromats in that colours that fall on, or near the same line of confusion, are likely to be confused by the respective anomalous trichromat. However, the greater distance between the two colours, the less likely individuals with a milder defect will confuse them.

25

Many colour vision tests select their colours based on the lines of confusion. Pseudo-isochromatic plates are an example. The figure and background on a given plate are isochromatic to a CVD observer when they fall on the same line of confusion.
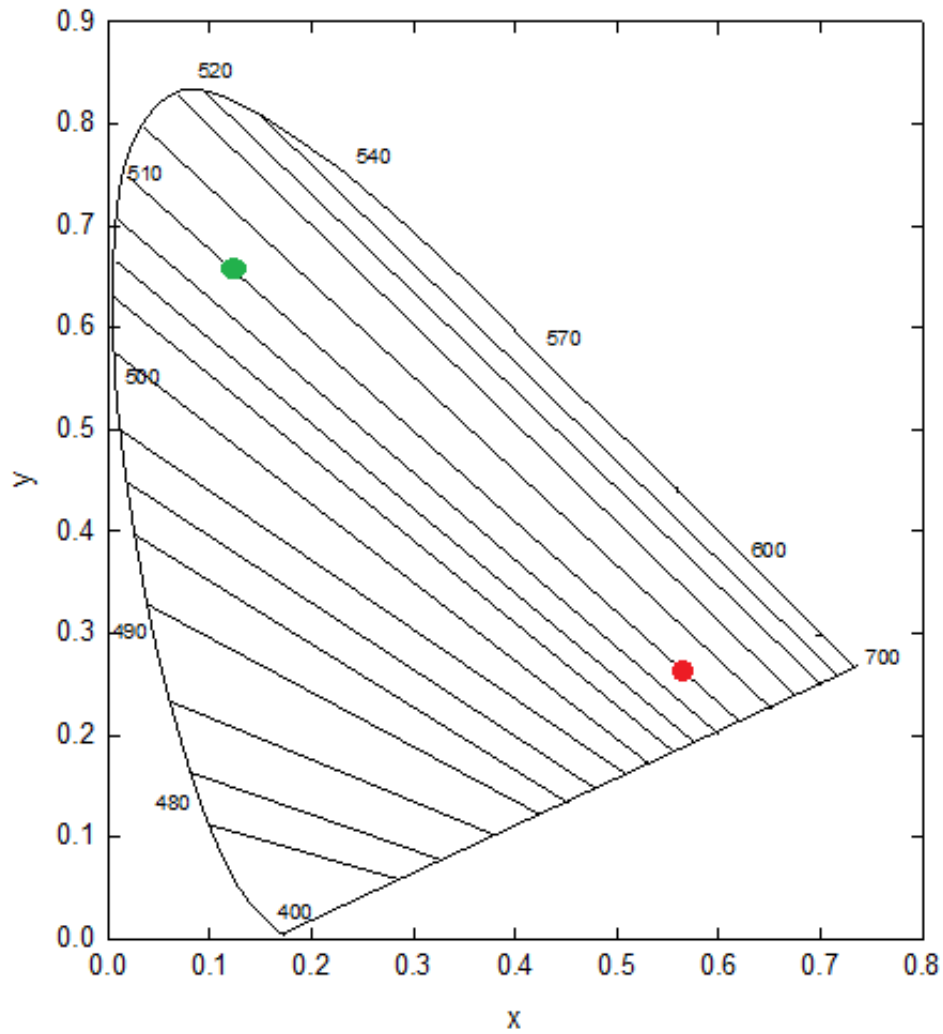
Figure 12. Deutan lines of confusion. The lines of confusion are based on the Vos and Walraven copunctual points, and the spacing between the lines is based on dichromatic wavelength

discrimination data from Pitt and Wright (Vos and Walraven, 1971; Walraven, 1974; Pitt, 1944; Wright, 1952)
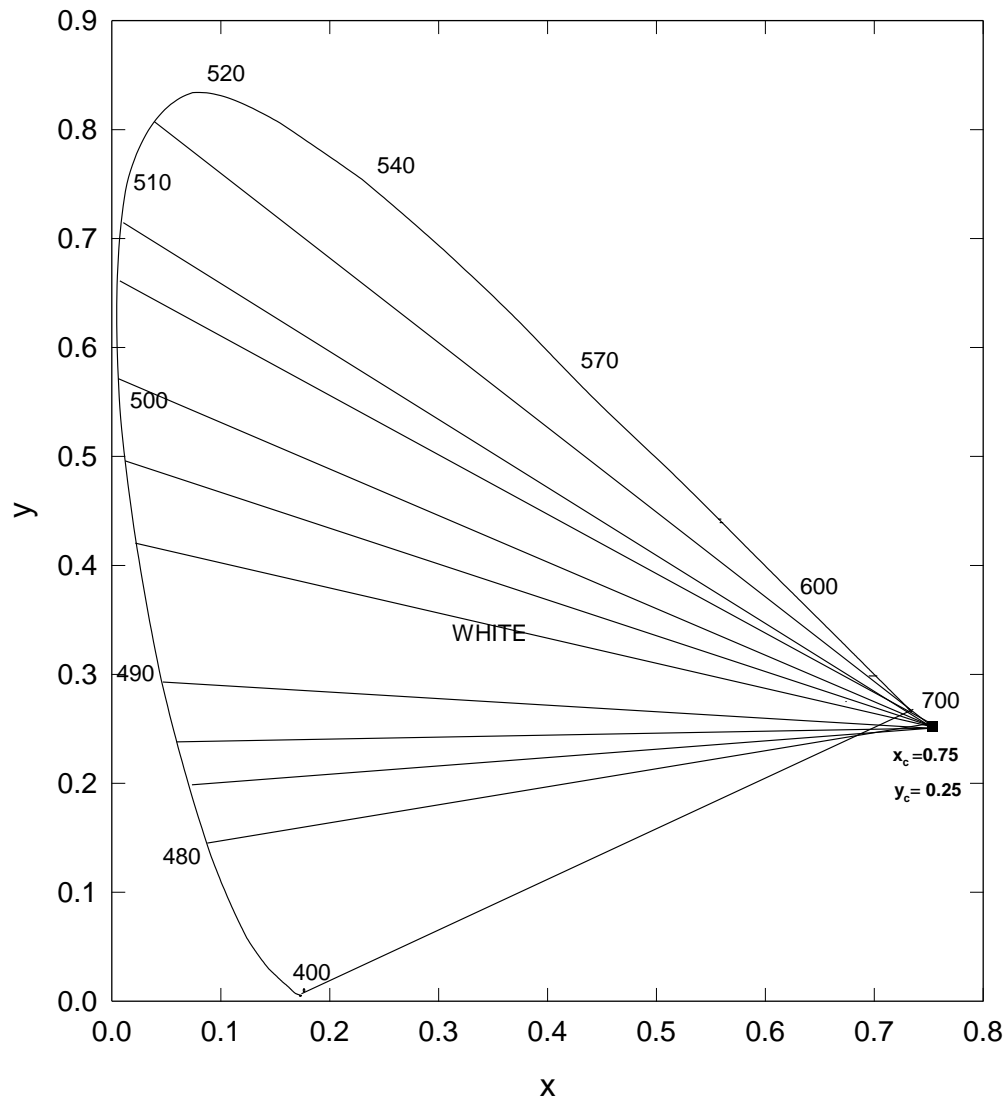
Figure 13. Protan lines of confusion. The lines of confusion are based on the Vos and Walraven copunctual points, and the spacing between the lines is based on dichromatic wavelength

discrimination data from Pitt and Wright (Vos and Walraven, 1971; Walraven, 1974; Pitt, 1944; Wright, 1952)
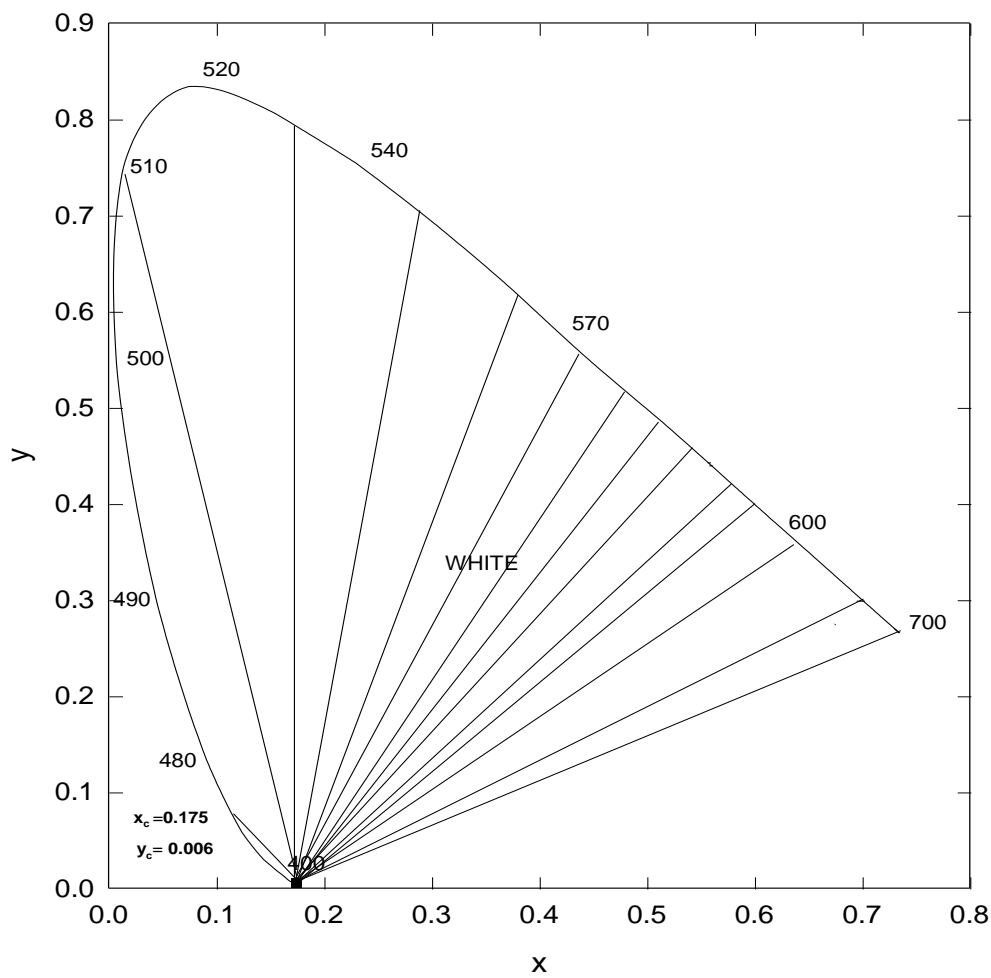
Figure 14. Tritan lines of confusion. The lines of confusion are based on the Vos and Walraven copunctual points, and the spacing between the lines is based on dichromatic wavelength

discrimination data from Pitt and Wright (Vos and Walraven, 1971; Walraven, 1974; Pitt, 1944; Wright, 1952)

1.3 Description and review of colour vision tests

The primary reason for testing colour vision in industries is that the individuals with congenital or acquired colour vision defects are at a greater risk of making errors in colour judgment. Today, there are a large number of colour vision tests to detect colour vision deficiencies and/or estimate one's ability to discriminate colours. The former include the various pseudoisochromatic plate tests, whereas the Farnsworth 100 hue and Farnsworth D15 tests are examples of the latter class of tests (Lakowski, 1966). Computerized colour vision tests are now becoming more common in the clinical setting. Different programs are available that screen for colour vision defects or do both screening and diagnose the severity of the colour vision defect. The computerized tests may also be easier to integrate with the person's electronic medical record.

Although there are a large number of tests available, the next section will describe only the tests in this study. The selection of tests was based on the current tests used in Canadian military and civilian aviation along with the Canadian railway industry, newer tests that are of interest to the Royal Canadian Air Force and tests that are under development.

1.3.1 HMC Oculus Anomaloscope

Although the anomaloscope is not often found in the clinical setting, it is considered the gold standard for determining the type and severity of a red-green defect. Figure 15 shows the

Oculus HMC anomaloscope. The Oculus HMC anomaloscope requires the person to carry out a Rayleigh colour match. For this match a patient adjusts the proportion of monochromatic red and green lights in a mixture so that the mixture matches the appearance of a reference monochromatic yellow light (Pokorny et al., 1979). The ratio of the red and green lights that match the yellow standard determines whether the person has normal or anomalous trichromatic red-green vision, and the range of acceptable matches determines the severity of the defect. Deuteranomalous individuals will require more green in the mixture with red to match yellow, whereas protanomalous individuals will require more red in the mixture. Their range of acceptable matches may be large, but anomalous trichromats do not accept all possible red-green settings as a match to the yellow reference. In contrast, red-green dichromats will accept all the possible mixtures of the red and green primaries as an acceptable match to yellow (Pokorny et al., 1979). However, some resolute dichromats who possess a single pigment gene will sometime not make a full range of match due to other factors such as optical density differences. Deuteranopes are distinguished from protanopes based on their brightness matches to the various mixtures of the red and green lights. Protanopes brightness match settings systematically decrease as the red-green ratio increases, whereas deuteranopes have a relatively constant brightness match settings for different red-green ratios.
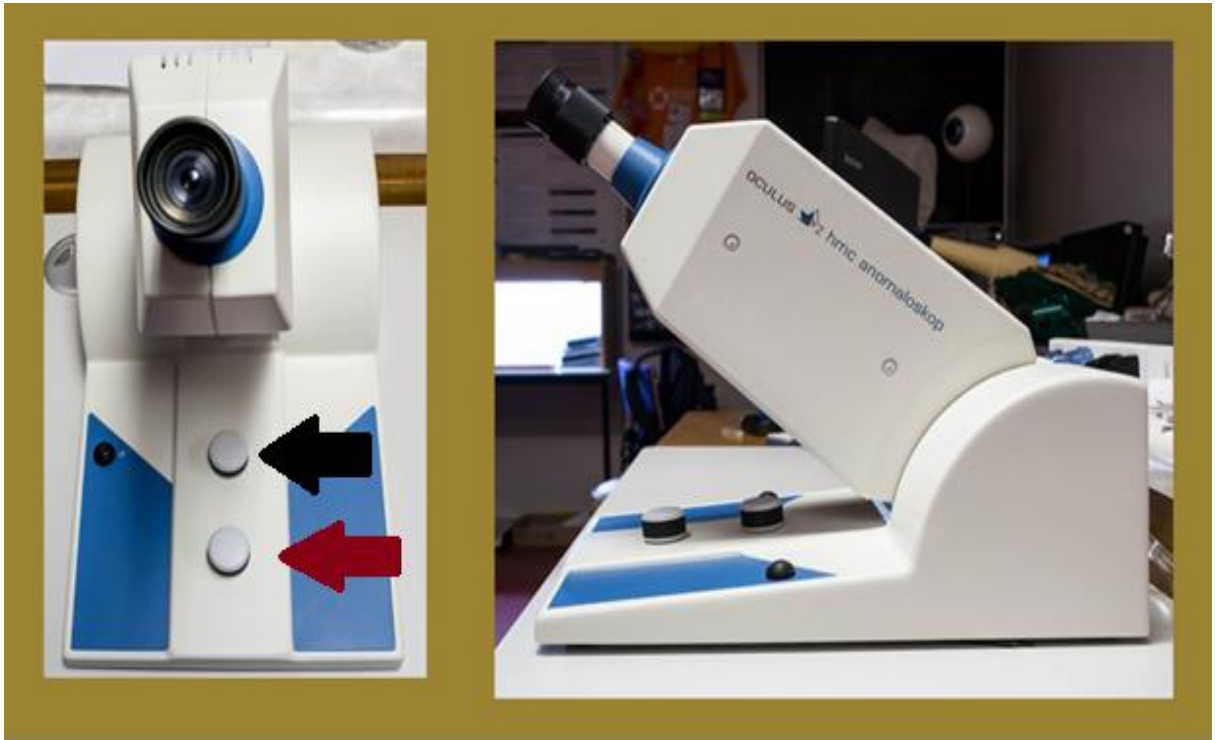
Figure 15. The Anomaloscope test. The participant views a circle of light through the eyepiece. The circle is divided into halves. The top varies in hue and brightness, and the bottom varies in brightness. Participants make a colour and brightness match by adjusting the two dials. The black arrow points toward the top dial that was used to control the colour of the top half of the circle. The red arrow points toward the bottom dial that was used to control the brightness of the bottom half of the circle.

Tritan, or blue-yellow defects, are also diagnosed based on colour matches. The one commercially available anomaloscope for diagnosing tritan uses the Moreland equation. The subject is required to adjust the relative amounts of blue and green monochromatic lights in a mixture to match a desaturated blue-green (cyan) reference. The blue and green primaries were selected to minimize inter-subject variability in their matches due to differences in their

macular pigment density (Moreland and Kerr, 1979). The reduction in the between-subject variability would reduce the number of false positives that could occur due to a higher macular pigment density. However, the colours used in the Moreland equation do not fall on a tritanopic line of confusion, and so the Moreland equation is limited in its ability to distinguish between tritanomaly and tritanopia (Moreland and Kerr, 1979). Although a tritanope may exhibit an extended matching range, they may not accept the entire range of settings as a match, which would be characteristic of dichromat colour vision.

1.3.2 Pseudoisochromatic Tests

One of the most common formats used in colour vision screening tests is the pseudoisochromatic format. This format uses a figure of one colour within a background of another colour. The background and figure colours are selected so that they appear identical to a person with a congenital colour vision deficiency. There are variations in brightness, saturation and hue within the figure and background colours. This "noise" helps to ensure that the person identifies the figure based on their ability to distinguish the dominant hue of the figure from the dominant hue of the background and not differences in brightness.

The pseudoisochromatic figure designs are vanishing, transformation, hidden or diagnostic. Vanishing plates are designed so that the colour-normal sees a figure and the colour-defective does not report any figure. Transformation plates are designed so that a colour-normal reports one number and colour-defective reports a different number. The hidden plates are the opposite of the vanishing plates. For these plates, the colour-normals

should not report any figure, whereas the colour-defects do report a number. Diagnostic plates are usually presented to only those who failed the screening portion of the test. These plates are a vanishing design used to identify a red-green defect as either protan or deutan. There are usually two different coloured figures on a plate. Protans should miss one coloured figure and deutans should miss the other figure. The diagnostic plates may also estimate the severity of the defect by systematically varying the average colour difference between the figure and background.

1.3.2.1 Ishihara Test

This test is the most widely used colour vision test for detecting red-green deficiencies; however, it does not screen for blue-yellow defects. The 38-plate edition is considered the gold standard for red-green colour vision screening (CIE, 2001). This edition has 25 plates with numerals of one colour embedded in a background of a different colour and 13 plates that have a path between two XXs defined by one colour within a background of another colour (Birch, 1997). These latter plates are designed for individuals who are unfamiliar with numbers. The numeric plates are divided into demonstration (plate no.1), transformation (2-9), vanishing (10-17), hidden digit (18-21), and diagnostic (22-25). Figure 16 shows the Ishihara test booklet and demonstration plate.
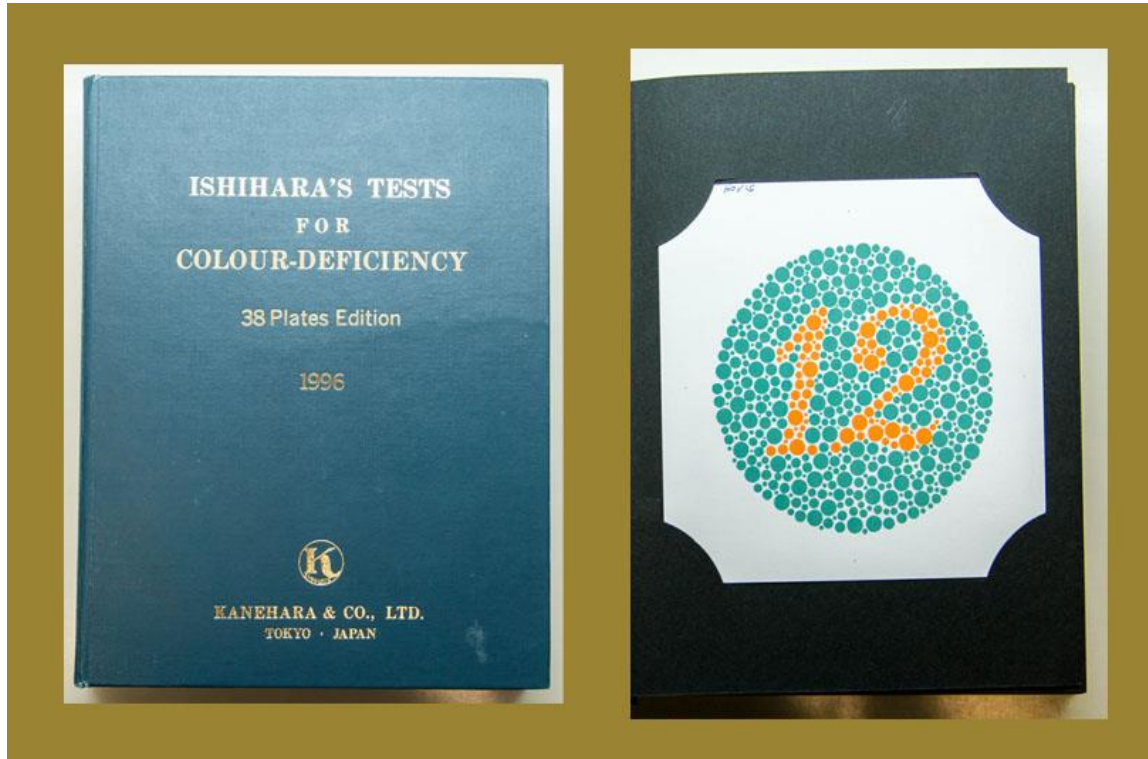
Figure 16. The Ishihara test booklet and an example of the demonstration plate on the right.

The test has a high sensitivity and specificity with near perfect agreement with the Rayleigh match as to whether the person is colour-normal or colour-defective (NRC, 1981; Birch and McKeever, 1993; Birch, 1997). Nevertheless, there is general agreement that the hidden plates are inefficient in terms of screening and the responses to these plates can usually be ignored (Belcher et al., 1958; Haskett and Hovis, 1987; Birch, 1997; Rodriguez-Carmona et al., 2012). Birch recommended counting errors on just the transformation and

vanishing plates (i.e. the first 17 plates) and use a failure criterion of 4, or more errors. Using this testing protocol, the sensitivity and specificity are 98.7% and 94.1%, respectively.

The diagnostic plates attempt to diagnose the defect as protan or deutan based on whether the person misses the pink figures (i.e., protan) or purple figures (i.e., deutan) in the grey background. One of the difficulties in interpreting these plates is that patients report that both the protan and deutan test figures are missing or that both of the figures are present. Birch reported that 83.2% of protans and 94.1% of deutans were classified correctly based on missing one of the two figures or identifying which figure was more distinct than the other if two numbers were reported (Birch, 1997).

1.3.2.2 Hardy, Rand Rittler 4th Edition Test (HRR)

This test screens for blue-yellow and red-green colour vision defects. It also attempts to grade the severity of the defect. It contains 24 plates that present either one or two coloured geometric shapes (X, Δ, and O) embedded in a background of grey dots. The test contains 4 demonstration plates, 6 screening plates (2 for tritan defects and 4 for red-green defects) and 14 plates designed to determine the type and severity of a red-green defect or the severity of a tritan defect. Figure 17 shows the HRR booklet and one of the demonstration plates. The type of red-green defect is determined by whether the patient misses the protan or deutan figures. The saturation of the coloured figures on the diagnostic plates increases as one proceeds so that individuals with a milder defect only miss figures on the initial diagnostic plates and individuals with a severe defect miss nearly all the diagnostic figures that

35

correspond to their type of defect. The severity of the defect is graded qualitatively as mild, medium or strong (i.e. severe).
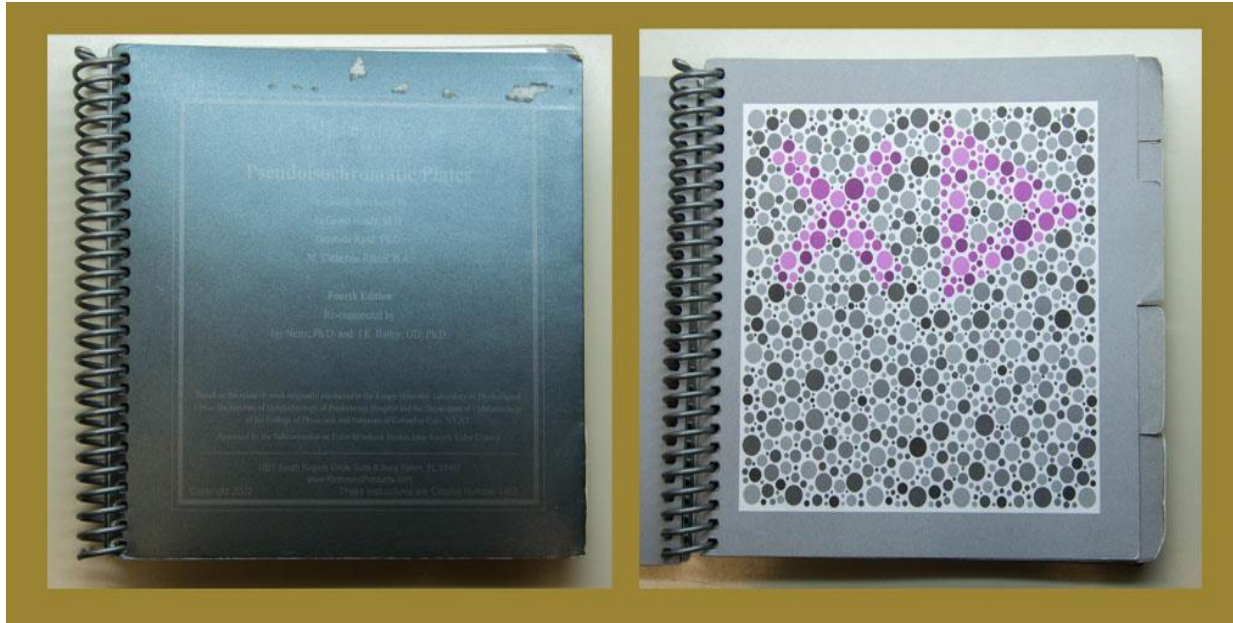
Figure 17. The HRR test booklet and an example of the demonstration plate on the right.

According to a study by Cole et al. (2006), the sensitivity and specificity of the HRR for red-green colour vision defects are 1.0 and 0.96 respectively when the failure criterion is two, or more errors, on the screening test figures. Eighty-six percent of the subjects were classified correctly as protan or deutan which was similar to the percentages reported by Birch for the Ishihara. The severity on the HRR was correlated with the severity diagnosed by the anomaloscope; however, 55% of individuals with a mild defect on the Rayleigh match (a range of matches less than 20 units) were classified as medium or strong on the HRR, and 37% of the dichromats were classified as medium(Cole et al., 2006). Nevertheless, Birch

does not consider the HRR 4th edition to have an adequate sensitivity (Birch, 2010). She

reported that the HRR had a lower sensitivity of 0.93 compared with Ishihara test, which had

a sensitivity of 0.98.

1.3.2.3 ColorDx Adult Pseudoisochromatic Plates

The ColorDx (PIP) test is a new computerized colour vision test that screens for red-green

and blue-yellow colour vision deficiencies, classifies the type of red-green defect, and

diagnoses the severity of the defect. The test plates are similar in design to the Ishihara

plates. Figure 18 shows an example of one of the screening plates.
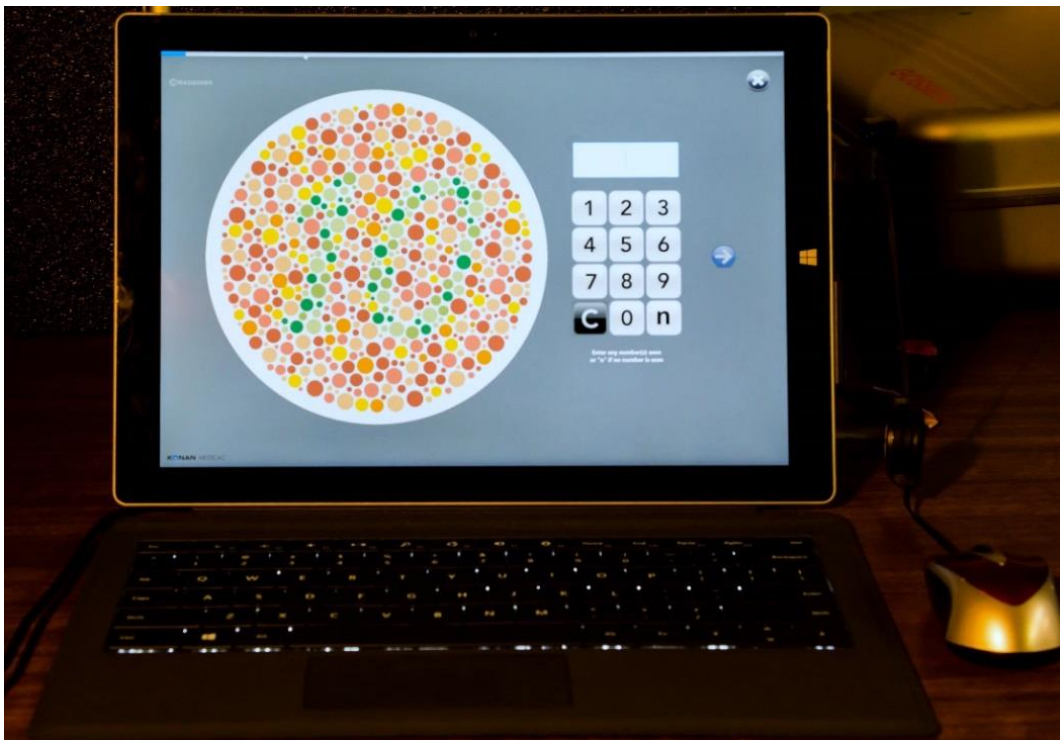
Figure 18. The ColorDx test showing a screening plate.

There are 25 images that are used to screen for red-green defects. These are followed by 12 tritan screening/classification plates. The red-green screening test ends once 5 errors are made, and the program switches to the blue-yellow series before proceeding to the red-green diagnostic series. The tritan plates classify the severity by presenting test figures, which vary systematically in their saturation. Individual who misses the more saturated colours is classified as having a more severe defect. The red-green diagnostic plates are administered next if the person failed the red-green screening plates. Half of the diagnostic plates have figure colours that a protan could miss and the other half have figures that a deutan could miss. Similar to the tritan plates, the saturation of the diagnostic figures changes systematically in order to classify the severity of the defect. The sensitivity and specificity of a prototype of the ColorDx were 0.96 and 0.99 respectively (Almustanyir & Hovis, 2015).

1.3.3 Threshold Tests

1.3.3.1 Cambridge colour vision test (CCT)

The CCT is one of the computer-based colour vision tests evaluated in this study. The test consists of a C shaped test figure that varies in hue and saturation relative to the reference grey background. The location of the gap in the C varies randomly from trial to trial in one of four directions: up, down, right, and left. The subject's task is to indicate the position of the gap. Figure 19 shows the Cambridge test and opening screen. Both the figure and

background colour vary randomly in luminance to ensure that only differences in hue are used to identify the target.
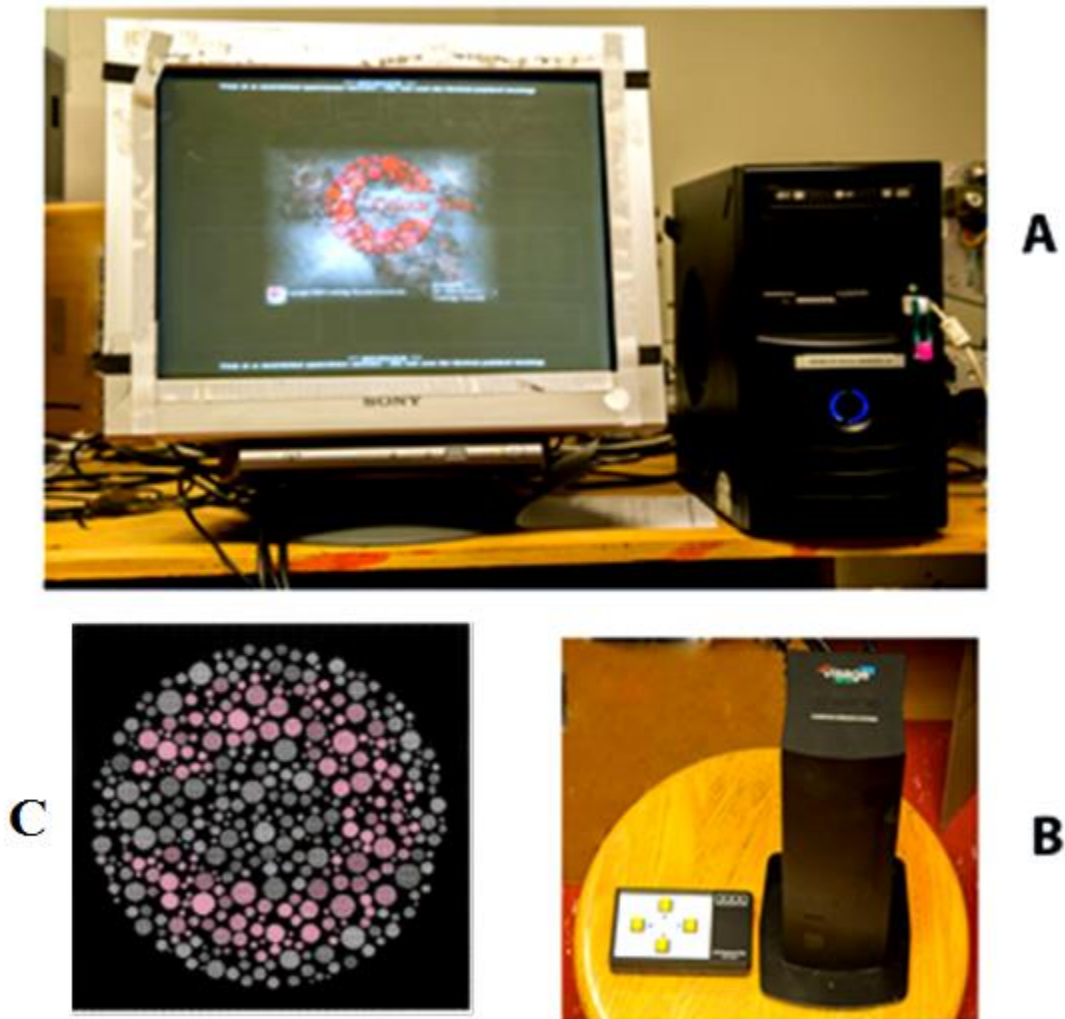
Figure 19. Picture of the Cambridge Colour Vision test. Section A is the display CRT monitor. Section B shows the response box and the graphics card tower that receives the subject's responses and generates the stimuli. Section C is an example of the stimulus.

There are two different testing protocols available. One is the Trivector and the other measures full discrimination ellipses. The Trivector test measures discrimination thresholds for three different colours from a grey reference. The colours were selected so that they fall on the each of the three-dichromat lines of confusion with the grey background. This test could be used to screen and estimate the severity of the defect. The full version measures chromatic thresholds for colours that are equally spaced around the grey reference. Thus, the full test determines general colour discrimination. The threshold data are fitted with an ellipse. The magnitude of the elliptical area is an index of a person's general chromatic discrimination ability, and the orientation of the ellipse shows the colour axis where their discrimination is the worst. Age-related norms for the ellipses have been published (Paramei and Oakley, 2014).

Different pass/fail criteria for the Trivector test have been evaluated (Regan et al., 1994; Shinomori et al, 2016). Nevertheless, using Shinomori, et al.'s data (2016) and the manufacturer's recommended criteria of more than 100 for the deutan and protan test colours, gives a sensitivity and specificity of 100% and 94% respectively for subjects between 18 and 60 yrs. The value of 100 represents a colour difference from the grey reference of 0.01 in the u'v' chromaticity diagram. Using cut-off score of greater than 150 resulted in slightly lower sensitivity and specificity values (Regan et al., 1994).

1.3.3.2 Rabin Cone Contrast Sensitivity Test (RCCT)

The principle of the RCCT is similar to the Trivector in that discrimination thresholds from a grey reference are estimated using colours that could be missed by each of the three types of colour vision deficiencies. The differences are that

i)   letters are used as the stimuli

ii)   No luminance or chromatic noise is present

iii)   Each colour is selected so that only one cone type is modulated as the saturation of the letter changes. That is, the test measures the discrimination threshold for the S-cone (i.e. detects a tritan defect), M-cone (i.e. detects a deutan defect), and L-cone (i.e. detects a protan defect) (Rabin, 2004). Figure 20 shows the RCCT.
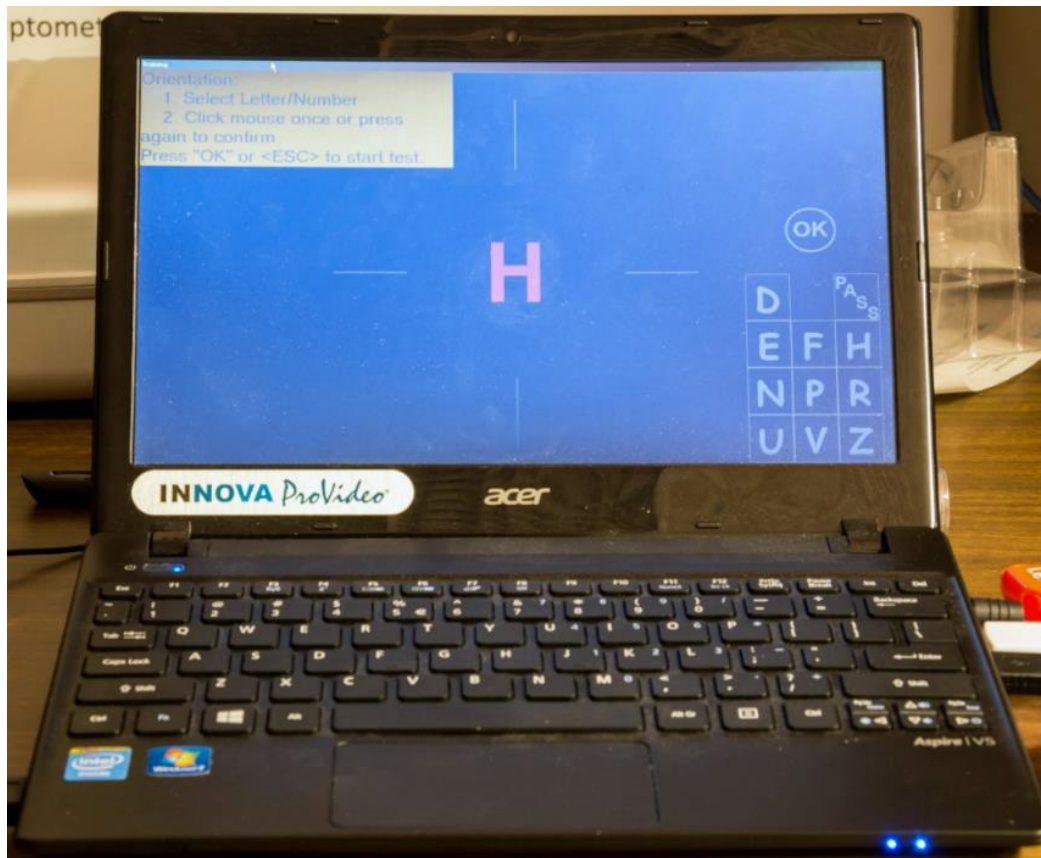
Figure 20. The cone contrast test with the demonstration target (letter H).

In Rabin's two studies (2004, 2011), the RCCT had a perfect agreement with the Rayleigh match with respect to whether the person had normal colour vision or a red-green defect. That is, the sensitivity and specificity were both 1.0. In a more recent study, Walsh et al., (2016) reported marginally lower values. The sensitivity of RCCT was 0.97 for both right and left eye, and the specificity was 0.97 for the right eye and 0.96 for the left eye.

1.3.3.3 Colour Assessment and Diagnosis (CAD) test

The CAD test can screen for colour vision deficiencies, measure chromatic discrimination around grey reference or both (Barbur et al., 1994, 2006). The CAD test consists of a grey background and coloured stimulus that is seen within a background of dynamic luminance contrast noise. Figure 21 shows the CAD test with a screenshot of coloured stimuli. The small individual squares making up the background and the stimulus change their luminance every 50 ms so that the display looks as if it is scintillating. As the individual squares oscillate in luminance, the coloured stimulus moves in one of the four diagonal directions (i.e., bottom left to top right, bottom right to top left, top left to bottom right, or top right to bottom-left). The subject task's is to press the appropriate button to indicate the corresponding direction of movement. A four alternative force-choice procedure is used to determine the observer's chromatic detection threshold in a specific direction within the CIE xy chromaticity diagram (Barbur et al., 2006; Rodriguez-Carmona et al., 2012). There are two general testing protocols available. One is fast screening and the other test measures thresholds for colours confused by protans, deutans and tritans. The data from this test could also be used to determine threshold ellipses around the grey reference. The test allows for different pass/fail criteria depending on the specific application of the results.
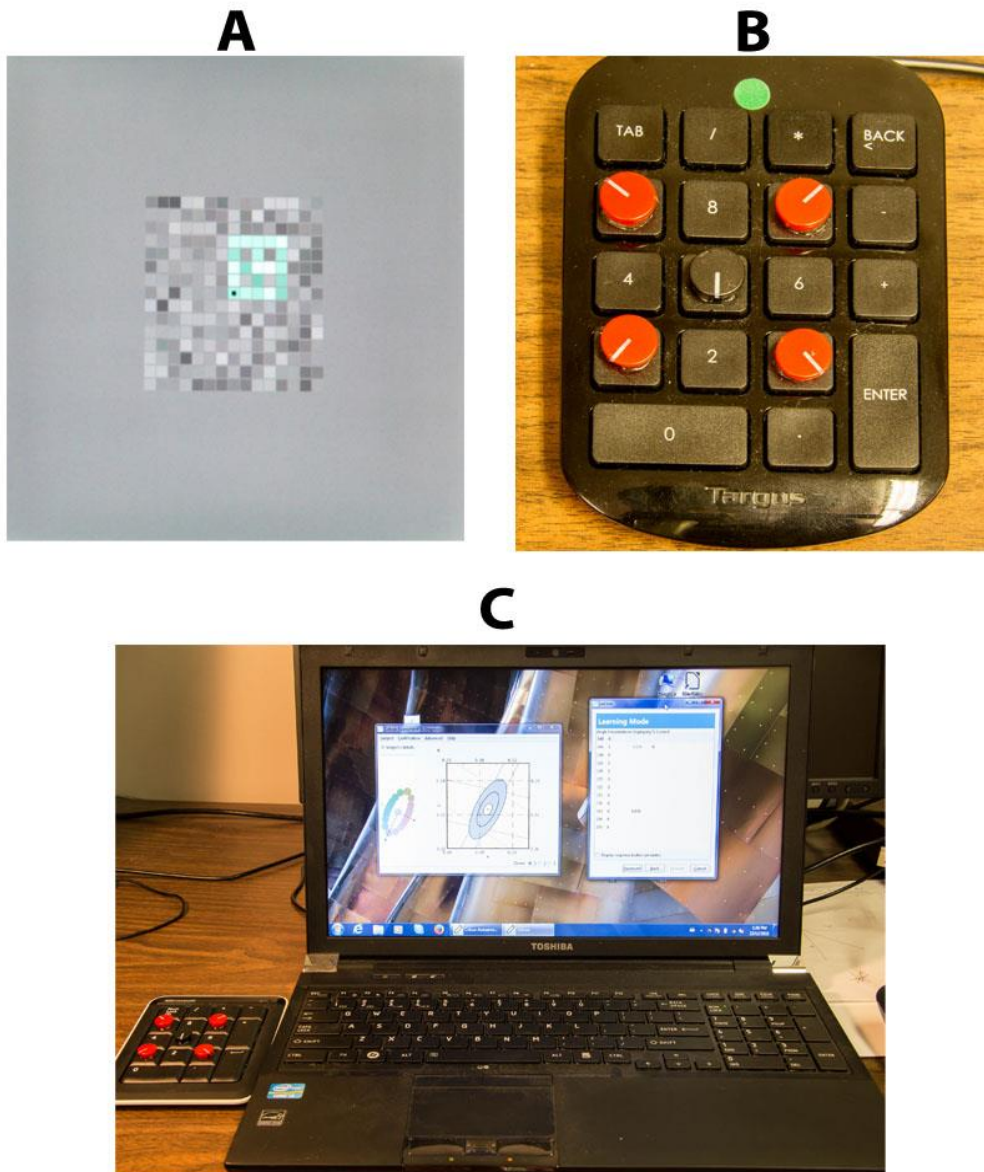
Figure 21. Picture of the CAD test. Section (A) presents the coloured square on the grey background (the test target). This target moves in four possible diagonal directions (bottom left to top right, bottom right to top left, top left to bottom right, and top right to bottom left). The

participant indicates where the target ends by pressing the corresponding button shown in B. Section (C) presents the laptop used to run the test.

The fast screening version screens for both red-green and blue-yellow colour vision defects. The moving square is presented at predetermined chromatic contrast for each of 16 different colours. The stimuli were selected to bracket the colours that would be confused with the grey background by the three different types of colour vision defects. There are 6 colours to screen for deutan defects, 6 colours to screen for protan defects and 4 colours to screen for tritan defects. The threshold test measures the chromatic thresholds for each of these colours. Values are given in units relative to the median value for colour-normals (referred to as Standard Normal Units or SNU units) and as the vector distance in the 1931 xy chromaticity diagram.

Figure 22 shows the normative data plotted in the 1931 xy CIE diagram from 238 normal observers (Barbur et al., 2006). The black cross in the center indicates the chromaticity coordinates of the grey background. The dark grey ellipse represents the median threshold value, or better, and the lighter grey ellipse denotes the ≤ 97.5 percentile region. Any threshold measurement that falls beyond the larger grey (97.5 percentile) ellipse would be abnormal. The blue, red, and green lines represent the test colours for the tritan, protan and deutan respectively.
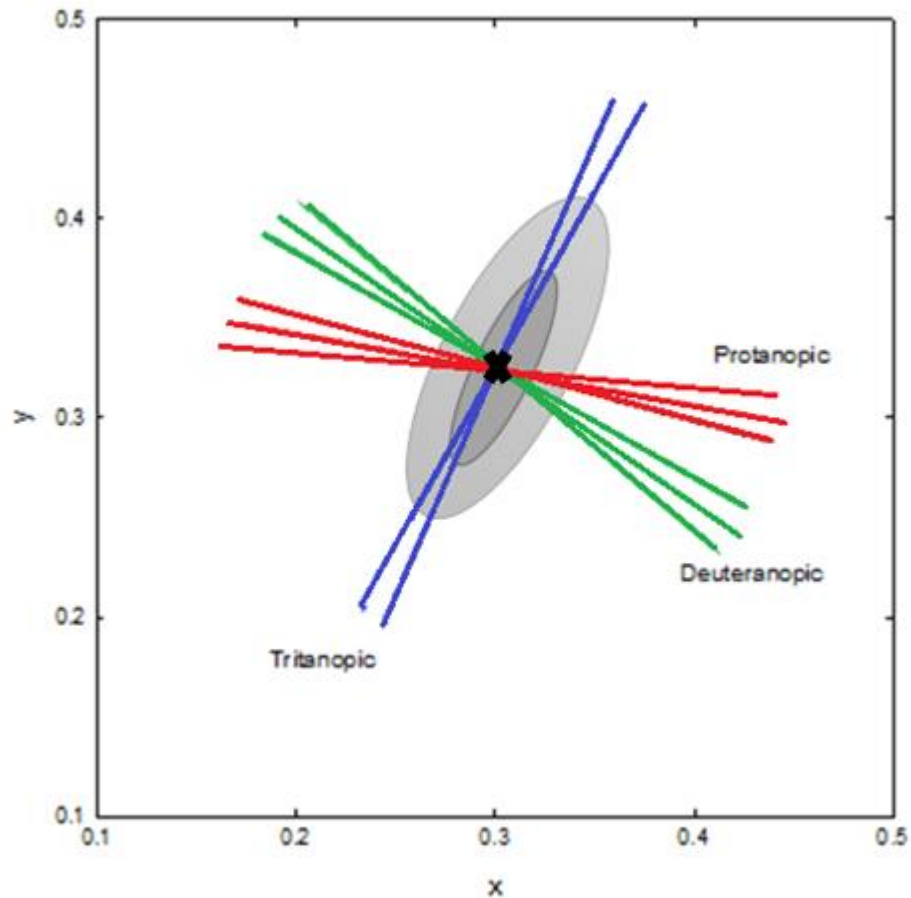
Figure 22. CAD threshold data plotted in the 1931 xy CIE diagram from 238 normal observers. The dark grey ellipse represents the values that are at the median or lower, and the light grey represents the thresholds $\leq$ 97.5 percentile of the colour-normals (Barbur et al., 2006). The green, red and blue lines are the test colours for the deutan, protan, and tritan confusion lines respectively.

Barbur et al. (2006) used the CAD test to assess the colour vision of 250 individuals with a congenital red-green defect. The blue-yellow thresholds overlapped substantially with the colour-normal data, and these thresholds were not statistically significant. However, there

46

was a clear difference in the red-green thresholds. If a CAD score of >1.8 SNU was used to identify a person with a red-green defect, analysis of their Figure 3, showed that both the sensitivity and specificity of the test were equal to 1.0. Seshadri et al. (2005) also reported sensitivity and specificity values over 90% for the WEB based version. However, Walsh et al. (2016) reported specificity and sensitivity values near 0.85 for the CAD threshold test.

Hovis & Almustanyir (2017) have validated the screening portion of the CAD test monocularly using the most stringent failure criterion (air traffic controller protocol). Also, they have validated the threshold version binocularly. They found that the agreement with the Rayleigh match was only fair. This lower agreement was due to 40% of the CVN who failed the red- green portion (false positive). To reduce this false positive rate, they have recommended testing those subjects who failed the screening portion with threshold test. The sensitivity and specificity of the screening portion were 0.98 and 0.60 respectively. The repeatability of the CAD screening was very good with 90% agreement between the first and second visit. However, the agreement with the Rayleigh match in term of screening for red-green colour vision deficiency was excellent in the threshold mode. In addition, there was a high agreement between the first and second visit on threshold mode (Hovis and Almustanyir 2017).

1.3.3.4 Landolt C colour vision test (LandC)

One of the disadvantages of the RCCT is that the range of contrasts is limited so that the test cannot measure colour-normal thresholds. In order to provide this capability, the United States

Air Force (USAF) is developing an enhanced version of the RCCT. Figure 23 shows the

LandC colour vision test. The program displays Landolt Cs oriented up, down, left or right.

The subject has to indicate which of the four orientations is presented. Similar to the CAD, the

LandC may be used to screen for colour vision deficiencies or measure chromatic thresholds.

The screening program presents each of the three colours at a suprathreshold contrast multiple

times. More than one incorrect response on any letter would be a failure. The threshold

program varies the contrast of the C using the Ψ adaptive procedure to determine the

individual's thresholds (Kontsevich and Tyler, 1999).

Based on recent work, the screening mode had an only fair agreement with the

Rayleigh match. The sensitivity and specificity were 0.93 and 0.63 respectively (Hovis and

Almustanyir, 2017). The lower level of agreement was because of a 37% false positive rate

(i.e., colour-normals failed) on the red-green screening portion. The protocol recommends

that individuals who fail the screening portion should be evaluated with the threshold test to

determine whether the screening result was a true or false positive. The monocular and

binocular L-cone and M-cone threshold mode showed excellent agreement with the Rayleigh

match in screening for red-green colour vision defects. The repeatability data for the

screening mode showed that the proportion of subjects that passed the first session and failed

the second was higher relative to the proportion who did better on the second session.

However, for the threshold mode, the red-green threshold test pass/fail outcomes for both the

colour-normal and colour-defective groups were highly repeatable.
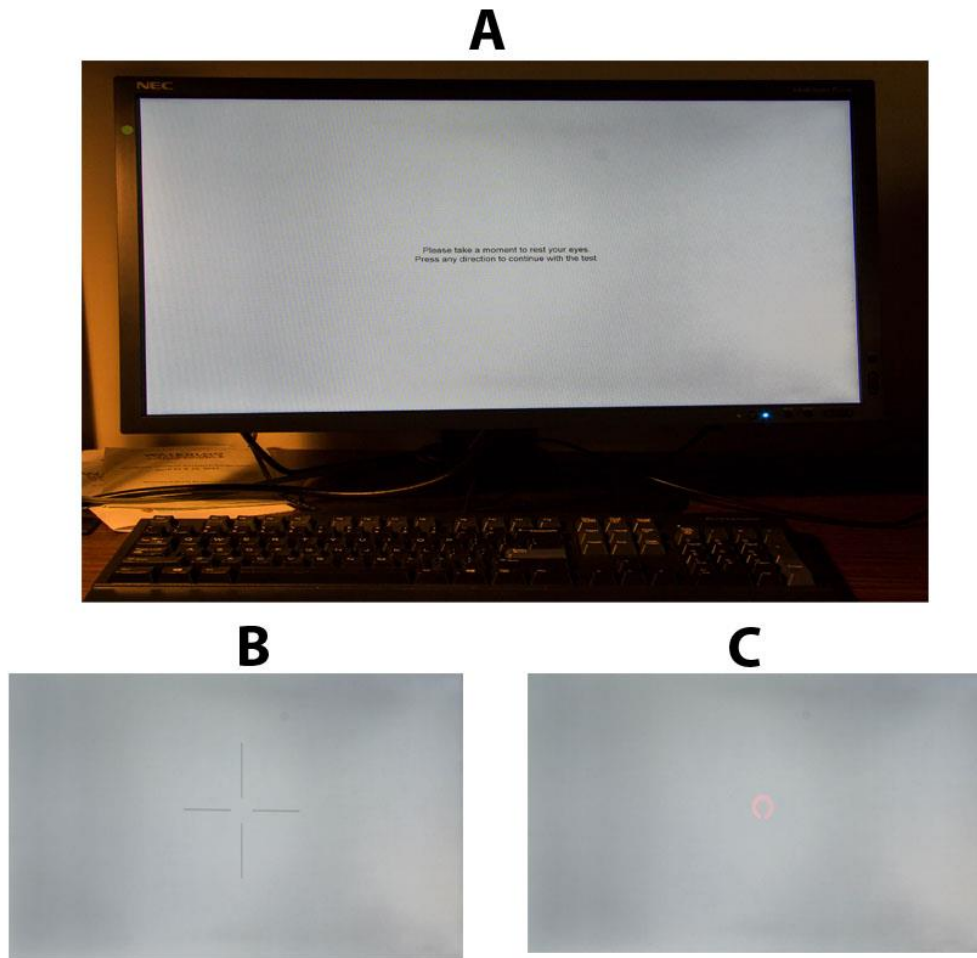
Figure 23. The Landolt C test. Section (A) the test monitor, section (B) the test field with the crosshair and in section (C) the letter C. The crosshairs help with fixation at the center of the display, but disappear when the stimulus is presented.

1.3.4 Colour arrangement tests

1.3.4.1 Farnsworth Munsell D15 (F-D15)

Farnsworth introduced the F-D15 colour vision test in order to distinguish colour-normal and those with a mild colour vision deficiency from those individuals with a moderate-to-severe colour vision deficiency (Farnsworth, 1947). Individuals who fail the F-D15 are more likely to encounter problems in making colour judgments in their everyday life or at work. The subject's task is to arrange coloured caps according to similarity by placing the colour sample that is most similar to the previous cap placed in the box. Figure 24 shows the F-D15 test with a CVN order and CVD order along with the score sheets. The numbers on the score sheet represent the cap numbers. Connecting lines are drawn between the caps in the order by which they were arranged. The connecting lines form a smooth curve if the caps are in perfect order and form a series of approximately parallel lines if there are major mistakes. Major mistakes are referred to as major crossings. The test is usually scored by visual inspection of the score sheet. Traditionally, 2 or more major crossings are a failure.

The F-D15 can also be scored using Colour Differences Vectors analyses (Vingrys and King-Smith, 1988). Three parameters are calculated: the Confusion index (C-index), Specificity index (S-index), and angle size. The C-index indicates the severity of the defect. It is correlated with the number of crossings and the total error score. The S-index provides a measurement of how regularly the crossings are oriented. A low S-index is an indication of a random arrangement. The angle gives a measurement of the type of the defect with protan angles larger than zero and deutan angles smaller than zero (Vingrys and King-Smith, 1988).
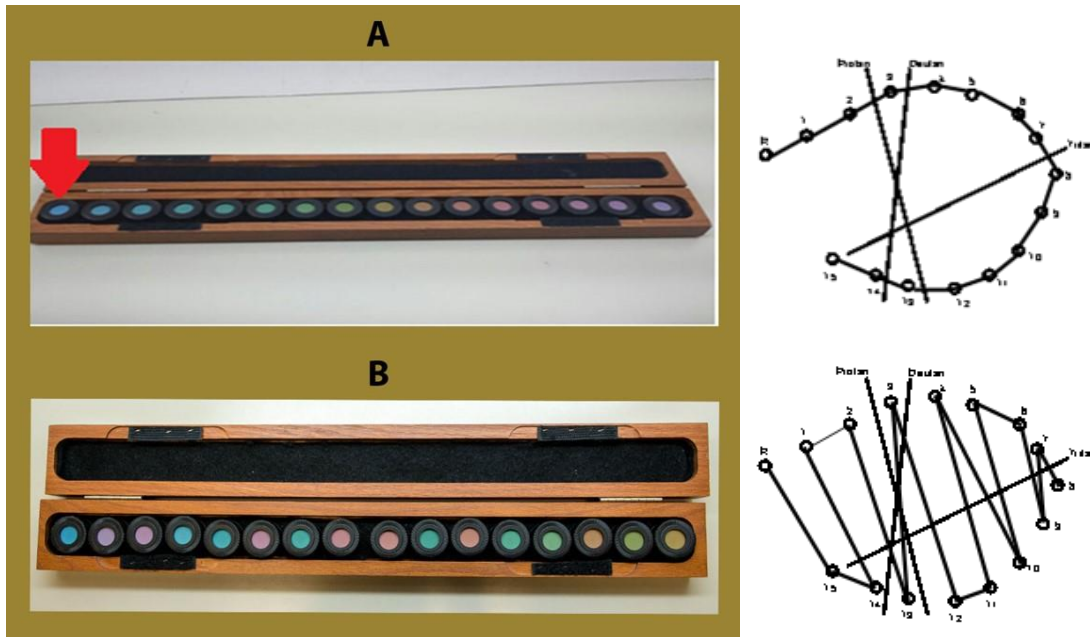
Figure 24. The Farnsworth-Munsell D15 colour test. (A) CVN F-D15 test arrangement and results drawn on the sheet. The red arrow indicates the reference cap. (B) F-D15 test with arrangement of a subject with a moderate to severe protan colour vision deficiency and the corresponding score sheet.

The pass rate for colour-defectives ranges from 45% to 53% if one major crossing is allowed (Atchison et al., 1991; Hovis et al., 2004; Birch, 2008). However, it is possible for a small number of dichromats (i.e. 3%) to pass the F-D15 using this criterion. The percentage of dichromats passing reduced to 1.5% using any crossing as a failure (Birch, 2008). Hovis et al. (2004) determined the repeatability of the F-D15 using 116 red-green CVD subjects. They reported that if the failure criterion was 2 or more major crossings, the repeatability of the F-D15 was good with kappa (κ) coefficients of 0.84 but less than 0.96 value calculated from Farnsworth's data (Hovis et al., 2004). The reason for the difference was that Farnsworth

included a large number of colour-normals, which would improve the repeatability of the test because of colour-normals rarely, if ever, fail the test.

1.3.4.2 ColorDx D15

The ColorDx program also has a computerized version of the F-D15. Figure 25 shows the test. The program requires the subject to drag the coloured circle up to the top of the screen in order to use that colour to "fill" one of the empty rectangles. The colour selected should be the one that is most similar to the previous rectangle filled. The colours in the rectangles may be rearranged. To our knowledge, no studies are reporting on its validity.
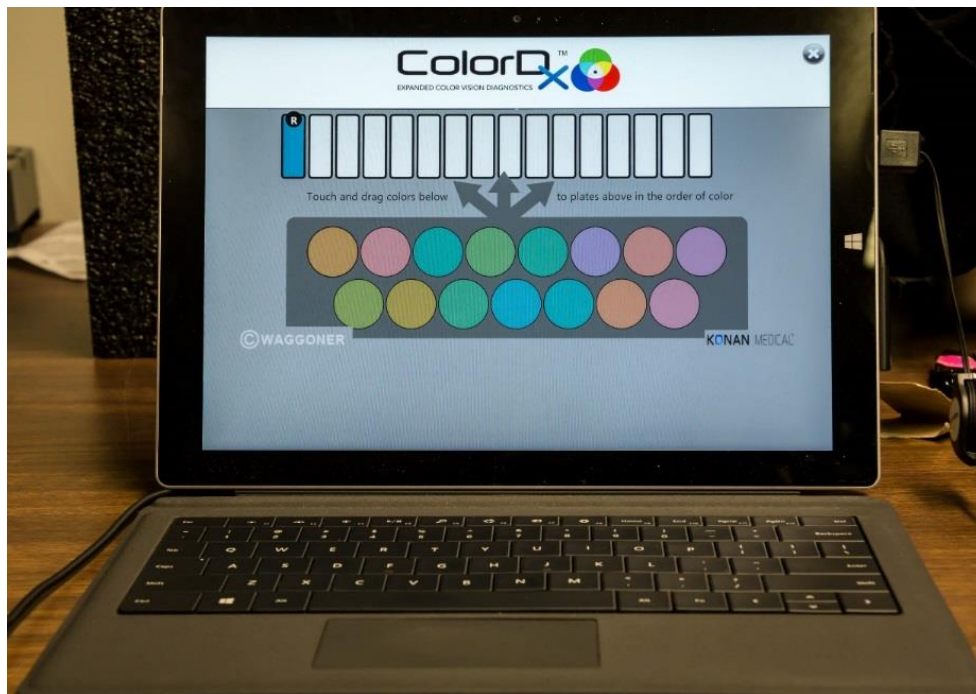
Figure 25. The ColorDx D15 test. The first blue rectangle on the top left is the reference rectangle.

1.3.5 Lantern tests

1.3.5.1 Holmes-Wright Type A

The Holmes-Wright Type A (HWA) colour vision test was designed to mimic navigation

lights used in the aviation industry (Holmes and Wright, 1982). Transport Canada Civilian

Aviation Medicine uses the HWA as an alternative to the F-D15. The test consists of three

lights (red, green and white) that are presented in pairs. The test colours fall within the

boundaries for aviation green, white and red. Thus, the test has face validity with the task,

unlike the Farnsworth Lantern, which does not use test colours corresponding to actual signal

colours (Cole and Vingrys, 1982). Figure 26 shows the Holmes-Wright Type A. The

individual lights subtend a visual angle of 0.9 min of arc at the recommended of 6 m. The

intensity of the lights at the corneal plane using the high brightness setting is nominally 5.4

μlux (Vingrys and Cole, 1983). There is some variation in brightness of the individual test

lights so that CVD cannot use brightness as a clue to identify the light's colour. Protans and

deutans tend to name the dim light colours red and green respectively (Hovis and Oliphant,

1998).

Figure 26. The Holmes-Wright Type A test.

Colour-defectives have difficulty passing this test (Vingrys and Cole, 1983; Birch, 2008b; Hovis, 2008). However, dichromats can pass the test if the passing criterion includes stopping the test if there is a perfect score on the first run of the nine pairs of lights. This is a problem because the dichromats (and other CVDs) cannot pass the test on repetition. The likely cause is too few presentation of the different test lights combined with some strategic guessing (Hovis, 2008). In order to avoid this problem, Hovis recommended that there should multiple runs of the nine pairs of lights and the failure score be based on the total number of

errors. He recommended using no more than 2 errors on 27 pairs to minimize the probability of a person passing the test the first session and failing the second. This criterion would also ensure that nearly all the colour-normals would pass on the first and second sessions.

1.3.6 CN lantern

Canadian railways use the CN Lantern to determine whether individuals with a colour vision defect can recognize railway signal lights correctly. The test consists of three lights (red, green and yellow) one above the other and a small black annulus surrounds it. The colours fall within the railway standards for wayside signal lights. The subject task is to name the colours displayed from top to bottom. There are 15 trials presented with three lights on each trial. Figure 27 shows the CN lantern test. When the test is viewed from 4.6 m, it is very challenging for individuals with a colour vision defect.

Figure 27. CN lantern test.

Hovis et al. (2006) determine the effect of the test distance on the CN lantern by using the standard 4.6m viewing distance and 2.3m (Hovis et al., 2006). There was a significant decrease in the mean number of errors from 7.6 to 4.3 as the test distant decreased. This reduction in the number of errors allowed more anomalous trichromat to pass the test. Only about 9 percent of the CVD passed at 4.6, but this improved to approximately 22% at 2.3m (Hovis et al., 2006). Changing the viewing distance, however, did not change the dichromat pass rate. Using the shorter test distance provides a method to decide whether the subject with colour vision deficiency could work safely in the railway yard where the signal lights are at shorter viewing distance than on the main track.

Chapter 2

Purpose

The thesis is divided into two projects. The first project was funded by Defence Research and Development Canada and the Canadian Institute for Military and Veteran Health Research. The first project will be presented in five chapters:

- ❖ Chromatic Discrimination Threshold Measurements for CVNs and Defectives: The purpose of this chapter is to transform the threshold data of colour-normals and CVDs into a common colour space in order to compare thresholds on a common scale. Then, all the tests will be compared to each other. The correlations with the Rayleigh and Moreland colour matches will also be determined.

- ❖ Colorimetric Analysis of the Farnsworth D15 and ColorDx D15 Color Vision Tests: The purpose of this chapter is to evaluate the Farnsworth D15 and ColorDx D15 colour vision tests in terms of their colorimetric properties. The order of the caps will be modelled using both normal and dichromat transformations of L*a*b* chromaticity space. Then, the predicted results will be compared with the arrangements made by subjects with a congenital red-green colour vision defect.

- ❖ Comparison of the ColorDx D15 with the Farnsworth D15 Color Vision Tests using two Scoring Criteria. The purposes of this study are to determine:

- o The repeatability within a session (1st trial vs 2 out 3 trials) of the ColorDx D15 and Farnsworth D15 with performing the tests without feedback using Farnsworth (1947) and Birch (2008) criteria.
- o The repeatability of between sessions for the first trial only and 2 out 3 trials using Birch's and Farnsworth's pass/fail criteria for both ColorDx D15 and Farnsworth D15.
- o The repeatability of classification for the 1st trial only and 2 out 3 trials using Farnsworth and Birch criteria for both ColorDx D15 and Farnsworth D15.
- o The validity of ColorDx D15 relative to the Farnsworth D15.
- ❖ Next Generation of Colour Vision Tests and Predicting the Occupational Tests: One of the factors to consider in revising the colour vision testing protocols is whether the newer colour tests could be adequate substitutes for the currently used tests. The ultimate goal is to determine which of the recent tests could be the best to replace the current occupational tests. This chapter will also provide a better understanding of how colour vision defects perform on the newer tests.
- ❖ Colour Vision Defectives Experience with Colours: the study will administer a short questionnaire to participants to determine whether individuals with a colour vision defect reported similar problems regarding colour-related tasks reported in the earlier surveys.

Saudi Arabia Cultural Bureau and King Saud University funded the second project. The second project is divided into two chapters:

- ❖ The CN Lantern Test and Different Viewing Distances: The first purpose of this chapter is to further quantify how viewing distance influences the error rate on the CN Lantern test. The work expands on a previous study, which looked at the pass rate for 4.6 m and 2.3 m test distances. This study will look at the pass rate for 4.6, 2.3, 1.15, 0.57m. The second purpose of this study is to determine the repeatability of the lantern results at the various distances in terms of passing or failing.

- ❖ Predicting the CN Lantern Test for Railways with clinical colour vision tests: The purpose of this chapter is to determine how well various clinical colour vision tests could predict CN lantern results at various test distances.

Chapter 3

Subjects and testing protocol

3.1 First Study

The first study recruited 60 CVN subjects and 68 subjects with congenital red-green CVD through posters, social media, posters on buses and newsletter advertisements. Appendix 5 shows the posters used in this study. The CVN participants were 50% females and 50% males whereas the colour abnormal group were predominantly male (89.7% males and 10.3% females) because it is X linked-recessive- trait. The subjects ranged in age between 17 years and 60 years and had no known vision problems other than a colour vision problem or a corrected refractive error. The 17 to 60 years age range was set primarily because of the Canadian Air Force age requirements. The median age for the CVN was 23, and for the CVD group was 24. Colour vision was classified according to the Rayleigh colour match using the manual mode on the HMC Oculus anomaloscope (Oculus Optikgeräte GmbH Wetzlar, Germany). There were 8 deuteranopes, 32 deuteranomalous, 19 protanopes and 9 protanomalous. One of the deuteranomalous individuals was classified as normal by the anomaloscope, but failed all the other colour vision tests with a deutan defect and so he was classified as a deuteranomalous. This case could be similar to the *pigmentfarbenambloyie* reviewed in Pokorny et al (1979). These individuals fail the screening plates but their Rayleigh matches results are normal. It is unlikely that this subject had just a figure-ground

60

problem in interpreting the pseudoisochromatic plate tests because as the subject also failed the D15, LandC and lantern tests.

Ocular diseases were ruled out using a short questionnaire. The possibility of a bilateral disorder associated with acquired colour vision defect was reduced further by restricting the subject pool to only those with a monocular visual acuity of at least 6/6 in better eye and 6/9 in the other eye at 6 m. In order to ensure that they had an adequate intermediate vision, they also had to have 6/24 in better eye and 6/30 on the other eye at 100 cm, and 6/12 in better eye and 6/15 on the other eye at 40 cm with or without spectacles or contact lenses. Tinted contact lenses or spectacles were not allowed. Subjects were asked to return within approximately 10 to 15 days to repeat the tests. Ninety-three percent of the colour-normals and 86% of the colour-defectives participated in both sessions. This study received ethics clearance through the Office of Research Ethics, at the University of Waterloo (Reference number: ORE 20996).

3.1.1 Testing sequence

The study started with an explanation of the study and obtained informed consent. Next, the subject filled in a questionnaire to determine whether the person met the inclusion criteria. Individuals who reported problems with colours were asked additional questions about their colour vision. Monocular visual acuity was measured from 6 m using a Bailey-Lovie chart and at 100 and 40 cm using a Reduced Snellen letter acuity chart. The order of the colour

vision tests was determined using a random block design. The tests were administered in the reverse order at the second session.

3.2 Second study (Railway studies)

The study recruited 56 CVN subjects and 63 CVD subjects. Appendix 6 shows the poster used in this study. All these subjects participated in the first study. The CVN participants were 50% females and 50% males, whereas the colour-defective group was predominantly male (90.5% males and 9.5% females). Based on the HMC Oculus anomaloscope, there were 7 deuteranopes (11%), 28 deuteranomalous (44.5%), 19 protanopes (30%), and 9 protanomalous (14.5%) in the CVD group. One of the deuteranomalous individuals was classified as normal by the anomaloscope but failed all the other colour vision tests with a deutan defect, and so was classified as a deuteranomalous. The median age for CVN was 23 yrs and 24 for the CVD. The subjects were asked to repeat the study after 10-15 days. Ninety-three percent of the CVNs and 94% of the CVDs did the second session.

Chapter 4

Material and Methods

4.1 First Study

4.1.1 Anomaloscope

The Oculus HMC anomaloscope tests for red-green defects using the Rayleigh colour matching equation and blue-yellow defects using the Moreland colour matching equation. The subject views the stimulus through the eyepiece. The stimulus consists of 2 adjacent semi circles. The top one displays the mixture of the two primaries and the ratio is controlled by the top dial. The bottom semicircle displays the reference stimulus, and the bottom knob controls its brightness. The experimenter can also control the anomaloscope settings through a computer interface. Figure 28 shows the computer display for the controlling the Rayleigh match and Figure 29 shows the control screen for the Moreland match.

The semicircle subtended 2 degrees for the Rayleigh match and Moreland equation. An additional eyepiece is added to the anomaloscope in order to the increase the size of the blue-yellow test field. The test was done monocularly starting with subjects preferred eye. For both tests, the stimulus was presented for 5 seconds followed by a white adaptation field of 3 sec duration in order to maintain a neutral adaptation state.
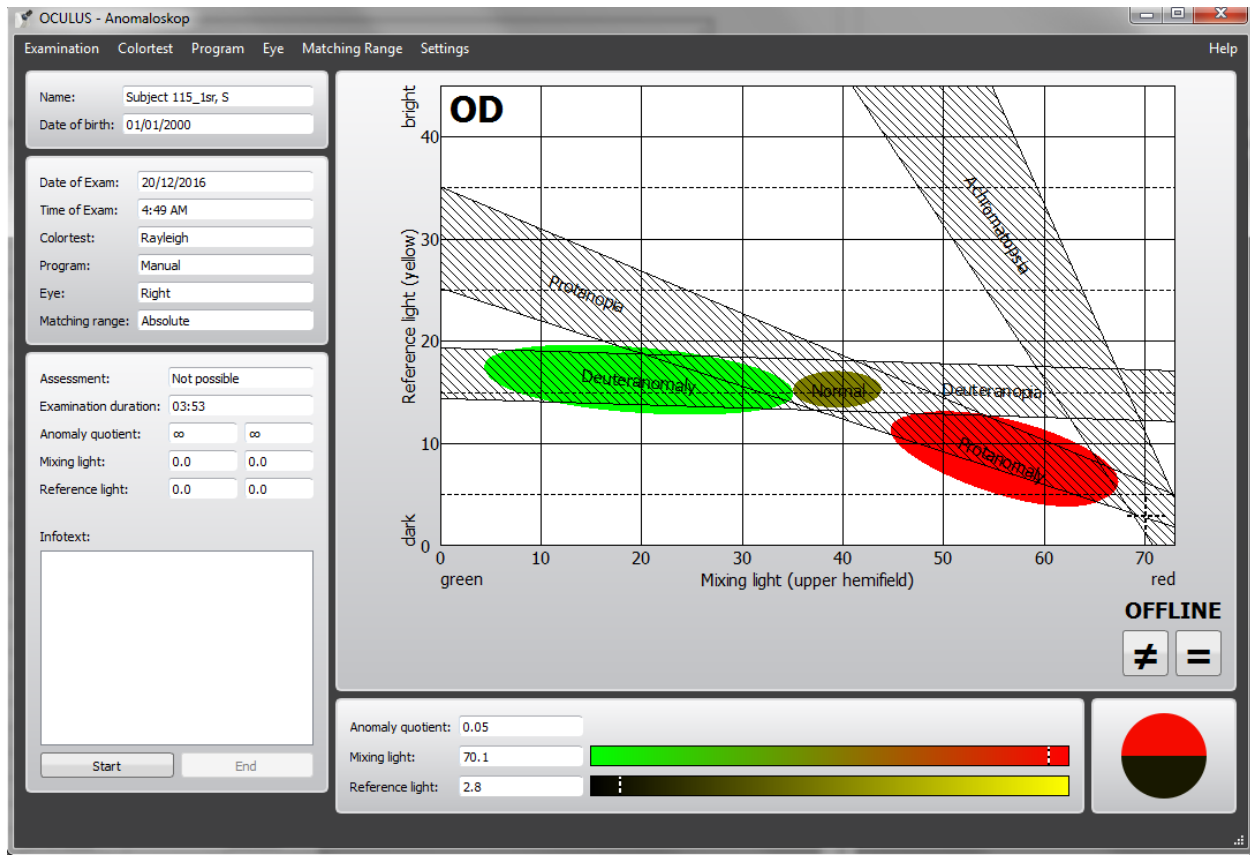
Figure 28. Screenshot for controlling window of Rayleigh test.

In principle, the subject's task is to adjust the relative amounts of the two primaries in the top half and the brightness of the bottom reference field so that the two halves of the circle look identical. This was not done in practice because the CVDs match tended to be highly variable. The actual procedure for Rayleigh match used was Linksz procedure for the Nagle anomaloscope (Pokorny et al. 1979). It was as following: The participant was presented with a typical CVN match. (For CVNs and some dichromats, the two fields could look identical). This presentation was used to familiarize the subjects with the stimulus and

task. Next, subjects were presented with different mixtures of the red and green primaries in the top field. The mixtures varied from 100% green (0 on the scale) to 100% red (70) on the scale in 10 unit steps. Subjects were asked to match the brightness of the top field by adjusting the brightness of the bottom reference light. After each brightness match, they were asked whether the top and bottom fields looked identical in both hue and brightness.

- There were three possible outcomes
- ➢ First, the subject reported that there was a colour match for all presentations (Full range of mixture settings was accepted). This result indicated that the subject was a red-green dichromat.
  - o The diagnosis of deuteranope or protanope was made by looking at the brightness match with the top field at 70 unit scale.
    - o Protanopes adjust the test light brightness value to a low value of less than 10 on the yellow brightness scale.
    - o Deuteranopes have brightness match setting that is within the range of CVNs (i.e., 15 to 25).
- ➢ Second, the subject matched at least one setting. The subjects were diagnosed as normal or anomalous trichromat.
  - o The diagnosis of normal vs. anomalous trichromat was made by determining the range of acceptable matches.
  - o The range of acceptable matches was determined by presenting the top field in steps within one unit scale below the match point for normals or below the lower

65

setting point of the acceptance for the anomalous trichromat. The subject adjusted the yellow light's brightness to make a brightness match and was then asked if the two halves looked identical. This process continues until the colours no longer matched. The upper limit of the range was then determined using the same procedure.

- For trichromats, the diagnosis depended on the actual range of acceptable matches.

  - If the range of acceptable matches was below 35 (more green) or the extent of the matches below 35 was greater than the matches in the normal or protanomalous range, then the person was diagnosed as deuteranomalous and the severity determined by the width of the acceptable match range.

  - If the range of acceptable matches was above 45 (more red) or the extent of the matches above 45 was greater than the range of matches in the normal or deuteranomalous range, then the person was diagnosed as protanomalous, and the severity was determined on the width of the acceptance match range.

  - If the range was between 35 and 45 or the extent of the matches was greatest in this range, then the person was diagnosed with a normal trichromat

➢ Third, if the subject could not make a match, then the subject task was to adjust the red-green values to make a match to the yellow reference. Next, match the brightness of the yellow field to match the top field and repeat until the two fields looked identical.

  o If the match was in the normals range (40±7), subjects were diagnosed as normal.

  o If the acceptance match was in the green area (0-35 unit scale) or in the red area (45-70 scale unit) the subject was diagnosed as anomalous trichromat (either deuteranomalous or protanomalous respectively).

  o The acceptance range was measured as described previously.

For the Moreland test, the experimenter has placed the target slightly outside the "Normal" area outlined on the control screen. The subject's task was to adjust the top field to match the bottom field as closely as possible, and then adjust the brightness of the bottom to make a brightness match and repeat until the two fields looked nearly identical. There was never a colour match because the bottom field was always more saturated in appearance than the top field. This was repeated 4 times. Once the fourth match was made, the first match was discarded and the average of the remaining 3 was calculated.

The range of acceptable matches was found next by a bracketing technique. The examiner set the blue-green ratio either below or above the match point by 10 units. The subject was asked to adjust the brightness of the bottom field to match the top field and

report whether the two fields were an acceptable match. Depending on their response, the

blue-green ratio was set 5 units either closer or further away from colour match. This was

repeated in successively smaller steps as the experimenter converged on the range of

acceptable matches. The procedure was repeated to find the range of acceptable matches on

the opposite side of the match point.
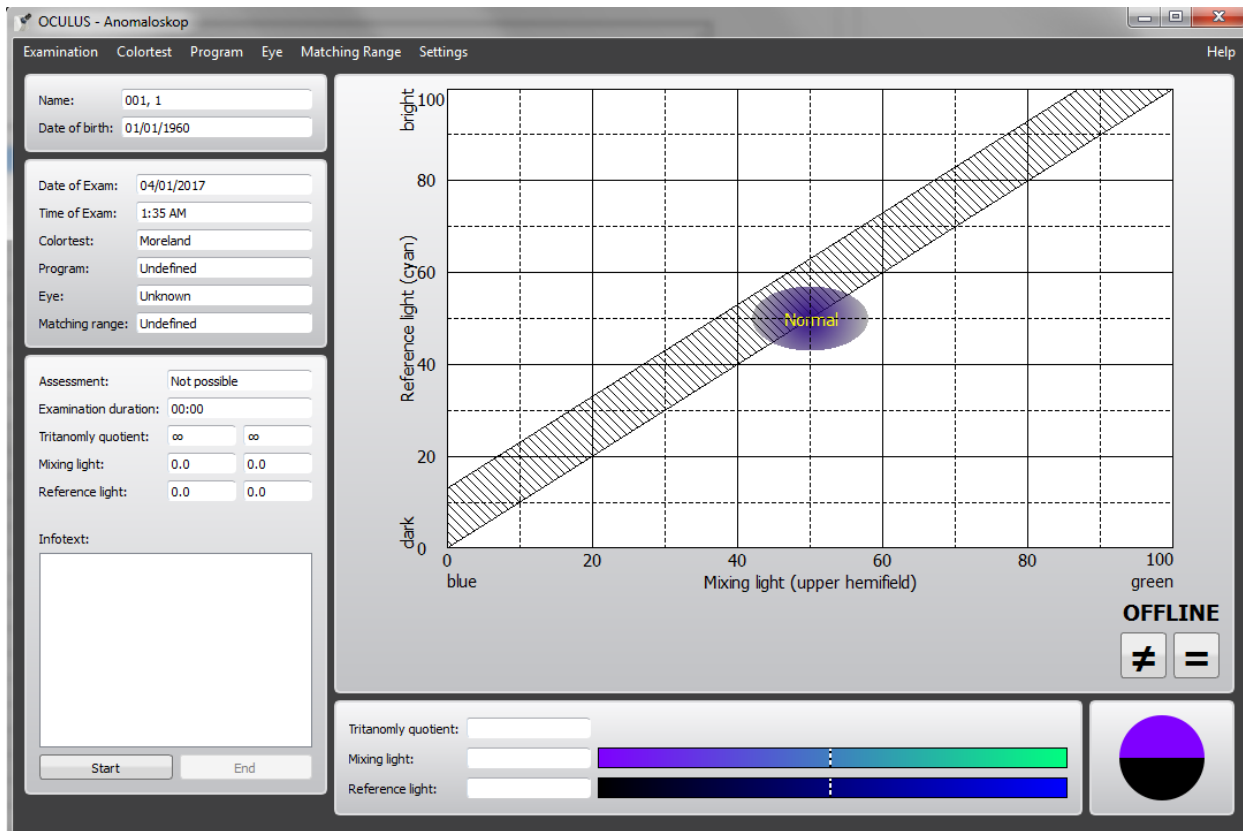
Figure 29. Screenshot for controlling window of Moreland test.

4.1.2 Pseudoisochromatic Plate Tests

All printed pseudoisochromatic plate tests were illuminated with an Illuminant C fluorescent lamp. The illuminance on the tests in the horizontal plane was measured at 1400 lux (±5%) using Minolta T-1 Illuminance Meter, Ramsey, NJ. All tests were viewed from approximately 60 cm. Subjects were allowed to view each page for approximately 5 sec before the next plate was presented. All pseudoisochromatic plates were viewed monocularly. Because the ColorDx PIP test scores each response as correct or incorrect, any response that was different from the normal response was recorded as an error on the other pseudoisochromatic plate tests. Table 1 summarizes the individual pseudoisochromatic plate tests failure and classification criteria. The next sections provide more information about the testing procedures.

4.1.2.1 Ishihara (38 plates)

The first 25 plates were administered to all subjects. This included the demonstration, transformation, vanishing, hidden, and diagnostic plates; however, pass/fail was based on the number of errors on the transformation, and vanishing figures and the diagnostic responses were only evaluated if the subject failed the screening series (Birch, 1997). Failure criterion on the screening plates was 3, or more errors, on plates 1-17 (Birch, 1997). If the subject reported two figures on the diagnostic plates, then s/he was asked which was more distinct. The classification was based on the column (protan vs. deutan) with the most errors or the majority of more distinct figures.

Table 1. Scoring criteria for the pseudoisochromatic plate tests.

Test	Failure Criterion	Classification Criterion for Red-Green Defects	Severity Criterion
Ishihara (38 plates)	≥3 errors on plates 1-17	Majority of errors or less distinct figures on the diagnostic plates in protan or deutan columns	NA
HRR 4th edition	**Blue-Yellow:** any error on screening plates **Red-Green:** > 1 error on screening plates; no errors on the diagnostic plates	Fewest number of errors in protan or deutan columns	**Red-Green** **Very mild**: Any red-green error on the screening plates and no errors on plates 11-20 **Mild**: Any error on plates 11-15 and no errors on the rest of diagnostic plates. **Moderate**: Any error on plates 11-18 and no errors on plates 19 and 20. **Severe**: Any error on plates 19-20 **Blue-Yellow** **Mild:** Any blue-yellow error on the screening plates and no errors on plates 21-24 **Moderate**: Any error on plates 21-22 and no errors on plates 23 and 24. **Severe**: Any error on plates 23-24
ColorDx PIP	≥3 errors on the blue-yellow plates or ≥5 errors on the red-green plates	Majority of errors on protan vs deutan diagnostic plates	**Red-Green** (defect with the majority of errors) **Mild**: ≥ 5, but less than 17 errors **Moderate:** ≥17 but <28 errors **Severe:** ≥28 errors **Blue-Yellow** **Mild:** ≥4 ,but less than 7 **Moderate:** ≥7, but less than 9 **Severe:** ≥9

4.1.2.2 HRR

Subjects were informed that there could be two, one or no geometric figures on each page. All the plates were presented to each subject. Any error on the blue-yellow screening plates indicated a blue-yellow defect and more than one error on the red-green screening plates and no errors on the red-green diagnostic plates were considered as normal (Cole et al., 2006). The classification was based on the HRR score sheet shown in Figure 30. The type of red-green defect was based on the column with the fewest errors. The severity of the defect was determined by how far the subject progressed through the diagnostic plates before there were no errors. An individual who made errors only on the red-green screening plates and none on plates 11- 20 was classified as very mild. Errors on plates 11-15, but none on the rest of the diagnostic plates resulted in a mild classification. A subject who made errors on plates 11-18, but none on plates 19 and 20 was be classified as having a medium, or moderate defect, and errors on plates 19-20 resulted in a severe classification.

Figure 30. Recording sheet for the HRR test.

4.1.2.3 ColorDx PIP

 Each plate was presented for 2 sec within a white background. After the number disappears, the subject enters the number that they saw or "N" if they did not see a figure using the keyboard or the touchscreen. They have 12 seconds to enter their response. The red-green screening plates are presented before the tritan test plates. If the subject failed the red-green screening portion, then the red-green diagnostic series starts. More than two errors on the

blue-yellow screening plates was considered as a blue-yellow defect, and more than four

error on the red-green screening plates indicated as a red-green defect. Table 1 shows the

criteria used by the ColorDx PIP to classify the severity of the colour vision defects. The type

of red-green defect is based on the maximum number of errors made on the protan and

deutan diagnostic plates. A Microsoft Surface Pro (Model number 1631) with Windows 10

operating system displayed the test. The viewing distance was approximately 40 cm away

from the subject in a dim room (1 lux). The monitor was calibrated using Spyder 4.5.4

colorimeter to a white reference of 6500° K correlated colour temperature. This calibration

was carried out every 30 days. The plate is presented within a white background with a

luminance of 90 cd/ m^2. The average luminance of the stimulus background is 46 cd/ m^2.

4.1.3 Threshold tests

4.1.3.1 Cambridge test

The CCT program (ver 2.3; Cambridge Research Systems, Ltd) was installed in a PC

computer (ASUS Intel Pentium 4 CPU 3.00 GHz) with Windows XP operating system. The

test figure was presented on a 21-inch CRT monitor (SONY model GDM-F520). The

monitor was calibrated using ColorCAL (Serial # 21001196; Konica Minolta CO. LTD)

colorimeter to a white reference of 6500° K correlated colour temperature every 30 days. The

subject viewed the monitor from a distance of 313 cm. At this distance, the Landolt C

subtended 4.3^o and the gap subtended 1.0^o. The luminance noise ranged from 8 to 18 cd/m^2.

The Landolt C was presented for 4 seconds. The Trivector test was performed first. The

discrimination ellipse was measured using 8 colours that were spaced equally around a grey background in the CIE u'v' chromaticity space. The center of the ellipse had the chromaticity coordinates in CIE u'v' of u'=0.197, v'=0.469. This chromaticity coordinate was the same one used for the reference point on Trivector mode. The room illumination was approximately 1 lux for both tests.

Subjects indicated the location of the gap using one of 4 buttons on a response box. If they could not see the ring or identify the location of the gap, then they were instructed not to respond and the next pattern will appear after a few seconds. Thresholds were measured by varying the average saturation of the C using a staircase procedure. The last 6 reversals were averaged to obtain the threshold value measured in the CIE u' v' chromaticity diagram. In the Trivector mode, thresholds that exceeded 100×10^{-3} u'v' units along the protan or deutan axis or 150×10^{-3} u'v' units along the tritan axis were considered as a failure. Classification of the defect was based on the highest threshold value. The pass/fail values for the elliptical parameters were determined using Receiver Operating Curve analysis described in later chapter.

4.1.3.2 Cone Contrast Test (RCCT)

The RCCT was the commercial version 16.02.0 supplied by INNOVA Systems (Burr Ridge, IL). It was displayed on an eleven inch Acer laptop (Model number: Q1VZC) using a Windows 7 operating system. The monitor was calibrated using a Spyder colorimeter (Express ver 4.5.4) to a white reference of 6500° K correlated colour temperature. This

version of the program required calibration on a weekly basis. The luminance of the grey background was 19.3 cd/m^2. The test was viewed from approximately 60 cm, which subtended 0.4 degrees. The test was presented in dim (1 lux) room lighting.

The test began by having the subject position themselves and/or the screen so that the screen was perpendicular to their line of sight. A 9 cm viewing tube that had a dot at each end was used to align the screen. Subjects viewed a target on the screen through the tube and adjusted their position and/or screen angle so that there is only one dot superimposed on the alignment target on the screen.

The program presents a single coloured letter in the centre of the screen for 5 seconds. The subject used a mouse to select their response from the key displayed on the screen. The test started with a practice trial to ensure the participant understood the task. The order of testing is L-cone sensitivity, M-cone sensitivity and S-cone sensitivity with the right eye tested before the left eye. Each sensitivity measure started with the highest cone contrast and used a staircase procedure to obtain the subject's maximum sensitivity. Cone contrast sensitivity was measured on a relative scale of 0 to 100 for each stimulus. One hundred is the highest cone sensitivity (lowest contrast), and zero indicates that the subject could not see the maximum contrast. A sensitivity that is less than 75 for any cone type is abnormal. The type of red-green defect was based on the minimum cone sensitivity.

75

4.1.3.3 CAD Test

The Z.X version of this test was installed on a Toshiba laptop (model number: TECRA R950-1EJ) with Pro Windows 8 operation system. The stimuli were presented on an NEC monitor (242 W-BK). The monitor was calibrated using LMT photometer (GOSSEN Germany) every 30 days. The average luminance of the grey background was 24 cd/m^2. At a viewing distance of 1.4 m, the side of the square stimulus subtended 1.6 degrees. The room illumination was dim (1 lux).

This test also has two options. The first is a screening test for both red-green and blue-yellow defects. The second classifies and quantifies the severity of the defect. The subject's task for both tests was to identify in which direction the square is travelling. There were 4 possible directions of travel in the diagonal direction. The subject was instructed to maintain fixation on the center of the square and not to track the moving target in order to obtain the best results. If they were unsure of the direction, then subjects were encouraged to make their best guess since a response was necessary to continue the test. If for any reason, they needed to repeat the same presentation, the experimenter would click on the "Represent" button in the measurement window. The air traffic controller protocol was selected. This protocol uses a very stringent pass/fail criterion.

The test began with a practice session with an easily noticeable stimulus to make sure that subjects understood the task. Next, the screening test was administered. This test presented suprathreshold (~2 SNU units) targets that could be confused with grey. In order to pass, the percent correct had to be greater than 66.67% in all directions. However, the

number of presentations for each direction varied according to the following algorithm. If the first three responses to a particular vector were correct, then the results were acceptable, and the presentation of that stimulus ended. If two errors were made, then the result is a failure for that colour, and the presentation of that stimulus ended. If one error was made within the first three presentations of an individual colour, then 3 more presentations were added to that individual colour. If the percentage correct was greater than 66.6% (i.e. only one mistake), then the result was a pass for that colour. If the percent correct was 66.6%, or less, in any colour, then the person failed the screening test. The program specifies whether the failure was red-green, blue-yellow or both.

Although the threshold program essentially measures a discrimination ellipse, the threshold vectors near the protan and deutan axes were averaged, as were the vectors near the blue-yellow axis. These averages are expressed in SNU units. The program also classifies the defect as protan, deutan or unclassified red-green based on the directions that have the highest thresholds. Receiver Operating Curve analysis was performed to determine the pass/fail values for the elliptical parameters which will be described in a later chapter.

4.1.3.4 Landolt C

The Landolt C program (ver 1.1.0) was run on a desktop (Lenovo Intel CORE i5) with a Windows 7 Professional Operation system. The stimulus was presented on an NEC monitor (Model 232 W-BK). The monitor was calibrated using X-Rite (Version EODIS3 i1) Display pro colorimeter every 30 days. The luminance of the grey background was 69 cd/m^2.

The current program is divided into a screening portion and a threshold portion. Ideally, if the person passed the screening portion, then there would be no need to measure the individual's thresholds. However, because the program is still under development, we measured the thresholds for each cone class regardless of the screening test results.

The Ψ adaptive threshold procedure allows one to measure just the threshold with a fixed slope of the psychometric function, or determine both the threshold and psychometric function slope (Kontsevich and Tyler, 1999). If the slope is fixed, the threshold variability asymptotes near twenty presentations, but there is a risk that the threshold could be biased. If the slope is also a free parameter, then the program estimates both the threshold and slope, but minimizing the variability of the threshold estimate has the highest priority in the initial presentations. Between 40 and 100 trials, the priority shifts to minimizing the variability of the slope estimate. After approximately 100, trials, it begins to minimize the variability of the threshold again (Kontsevich and Tyler, 1999). Based on the behavior of the adaptive procedure, the time required to run more than 30 trials for each cone contrast, and preliminary data from the USAF and our pilot studies, we elected to use a fixed slope of 2.6 for the L and M cone threshold, a fixed slope of 1.9 for the S-cone threshold and 20 presentations for each cone threshold for the monocular trials. Both the slope and threshold were determined using 30 trials for the binocular trials.

The subject viewed the test from 1m away from the screen, and their task was to press the keyboard direction corresponding to the direction of the gap in the C. At this distance, the Landolt C subtended 1.4 degrees, and the gap was 0.3 degrees. The room illumination was

dim (1 lux). The subject performed this test 3 times in the following order: right eye, left

eye, and binocularly. The order of presentations for all tests was randomized between L cone,

M-cone and the S-cone stimuli. Each monocular test began with the screening test. The

Landolt C had a fixed contrast of -1.66 log contrast for L cone and M cone and -0.55 log

contrast for S cone. These values are approximately 3 standard deviations from the USAF

colour-normal mean thresholds. There were 8 presentations of each contrast. Two, or more,

errors on any cone stimulus were a failure for the screening portion. Regardless of the

screening outcome, thresholds were measured after a short rest.

4.1.4 Arrangement Tests

4.1.4.1 Farnsworth Munsell D15

All the loose caps were removed from the box and arranged randomly on the table in front of

the subject. They were asked to place the coloured cap that is most similar to the previous

one placed in the box. They were allowed to rearrange the caps once they were placed in the

box. The test was administered three times without any feedback. The test was illuminated

with Illuminant C at 1400 lux (±5%). Scoring was based on both visual inspection and the

Color Difference Vector analysis (Vingrys and King-Smith, 1988). A major crossing was

defined as a difference between adjacent cap numbers that were greater than 2. Table 2 lists

pass/fail criterion for the Color Difference Vector analysis. The angular subtense of the caps was 1.5° at a 40 cm viewing distance.

4.1.4.2 ColorDx D15

The test was part of the ColorDx test suite. It was displayed using the same Surface Pro as the ColorDx PIP test. The individual test colours were presented as coloured disks in the middle of the screen. The viewing distance was 50 cm. The angular subtense of the rectangle was 1.15° by 2.9°, and the diameter of the circle was 2.3°. The average luminance of the rectangles and circles was 17 cd/m^2. The subject's task was to select the coloured disk that was most similar to the last filled rectangle at the top of the screen and drag it to the first empty rectangle. Table 2 lists pass/fail and classification criteria. The subject was allowed to rearrange the order. The ColorDx D15 was also administered 3 times without any feedback.

4.1.5 Lantern test

4.1.5.1 Holmes-Wright Type A

The test was viewed from 6m. The test was illuminated at 180 lux in a plane parallel to the floor at a table height (Holmes and Wright, 1982). The test started with the brightness set on DEMO and examples of red, green and white light were shown to the subject. Subjects could review the lights if they wished. The brightness was then changed to high, and 27 pairs of the test lights were presented. The starting positions for the second and third runs were randomly varied. There was no time limit for the presentation, but the subjects were encouraged to respond within 10 second. Table 2 lists the failure criterion.

Table 2. Scoring criteria for the D15's and Holmes Wright Lantern Tests.

Test	Failure Criterion	Classification Criterion for Red-Green Defects	Severity Criterion
Farnsworth Munsell D15	**Visual Inspection:** > one major crossing on two out of three trials **Colour Difference Vector :** C-index $\geq$1.78	**Visual Inspection:** Comparing error pattern to score sheet	NA
ColorDx D15	**Visual Inspection:** > one major crossing on two out of three trials **Colour Difference Vector :** C-index $\geq$1.78	**Visual Inspection:** Comparing error pattern to score sheet	NA
Holmes-Wright A	>2 errors on 3 runs (27 pairs)	NA	NA

4.2 Second Study

The colour vision tests used in this study were

A. Printed colour vision test

o HRR test: binocular testing.

o Ishihara: binocular testing.

B. Computer-based tests.

o CAD test threshold mode: binocular testing

o ColorDx (ColorDx PIP): binocular testing

o Rabin Cone Contrast Colour Vision test (RCCT): monocular testing

- o Cambridge Color Vision Test (CCT): binocular testing.

- o Landolt C cone contrast test (LandC): binocular testing.

C. Colour Sorting Tests

- o Farnsworth D15: binocular testing.

- o ColorDx D15: binocular testing.

D. Colour Matching Tests

- o Oculus HMC Anomaloscope Red/Green: monocular testing

All tests above were described previously in the previous section (first study).

E. Lantern tests
- o CN lantern

The CN lantern test presents 15 different triplets of lights. The triplets consist of three coloured lights (red, green and yellow) one above the other and they are each surrounded by a small black annulus. The first two presentations are the demonstration, and the next 13 are the test lights. The subject's task is to name the colours displayed from the top to bottom. A mistake in identifying any single light was counted as an error. The test was conducted at 4 different viewing distances (4.6m, 2.3 m, 1.15 m, and 0.57 m). At 4.6m, each colour light subtended 1.2arc min in diameter, which corresponds to approximately 500 m sighting distance for wayside signals. Halving the sighting distance will increase the angular size to 2.5, 5 and 10 min arc at 2.3m, 1.15m, and 0.57m respectively. These will be equivalent to viewing distances of 250 m at 2.3m, 125 m at 1.15m and 63m at 0.57m. The room illumination was 300 lux in a plane parallel to the floor at a table height. A single error was

allowed at the longest distance of 4.6 m providing it was not identifying a green light as red or vice versa. At the other distances, any error was considered as a failure. All of these tests were repeated in approximately 10 to 15 days.

Because all the subjects in this study participated in the first study, the only tests that they were required to do were,

1. CN Lantern Test,

2. Printed colour vision tests,

3. ColorDx PIP test.

The testing sequence started with the CN lantern at 4.6 m. Each subsequent CN lantern test distance was reduced by half. The pseudoisochromatic tests (Ishihara, HRR, and ColorDx PIP) were administered in between the lantern trials in order reduce any learning effects. The order of the pseudoisochromatic plates was determined using a random block design and the order of the pseudoisochromatic tests at the second visit was the reverse order of the first visit. The lantern test sequence was the same for both visits. This study received ethics clearance through the Office of Research Ethics, at the University of Waterloo (ORE 21094).

Chapter 5

Chromatic Discrimination Threshold Measurements for Colour Normals and Defectives

5.1 Introduction

Prior to the introduction of the computer-based colour vision tests, most of the clinical tests presented stimuli that were suprathreshold for CVNs and could only crudely measure chromatic thresholds of CVDs. Two exceptions were the Color Threshold Lantern and Gunkel Chromagraph (Rowland, 1943; Sloan, 1944; Gunkel and Cogan, 1978). Both were capable of measuring colour-normal thresholds, although neither test achieved widespread use. Most of the currently available computerized tests are capable of measuring chromatic thresholds. The Rabin Cone Contrast test is the exception. Although it can estimate thresholds for CVDs, it does not present stimuli with chromatic contrasts low enough to measure CVNs chromatic thresholds. All computerized tests measure chromatic thresholds from grey a reference background while the Cambridge Colour Vision test can measure chromatic thresholds against two other background colours. One of the drawbacks of these newer tests is that the thresholds are measured in different colour spaces, which makes it difficult to compare results across tests.

The purpose of this section is to transform the threshold data of CVNs and CVDs into a common colour space in order to compare thresholds on a common scale. The tests that will be compared are the Colour Assessment and Diagnosis (CAD), Cambridge Colour Test

(CCT) ellipses and Trivector (CCT Tri), Rabin Cone Contrast Test (RCCT), and the Landolt C colour vision test (LandC). The correlations with the Rayleigh and Moreland colour matches were also determined to find out whether any test could be a substitute for the anomaloscope, which is considered to be the gold standard.

5.2 Methods

In this study, the ellipse or vector length was calculated using the CIE 1931 as the common space. The reason for selecting this space is that many transportation industries still use the 1931 space to specific their colours and the 1931 space serves as the basis for all other chromaticity spaces. The parameters evaluated were,

1. Elliptical parameters of the CAD and CCT discrimination ellipse. Both tests were performed binocularly.

2. Vector lengths of the CCT Tri, LandC, and RCCT. All tests were administrated monocularlly.

5.2.1 Discrimination Ellipses

The CAD test measured chromatic thresholds in 16 directions. The reference was a grey background. The 16 vectors straddle the dichromat lines of confusion through the grey background. The 1931 chromaticity diagram is the native colour space. A four alternative force-choice procedure determined the observer's chromatic detection threshold in each direction. The program calculates the vector length for each of the 16 directions.

The CCT thresholds were measured in 8 different directions equally spaced around the grey background reference in its native u'v' chromaticity space. A staircase procedure was used, and the last 6 reversals were averaged to obtain the threshold. The test output included the grey u'v' coordinates (u_w, v_w) and vector length in each of the 8 directions. To convert the results into xy space, the following calculations were done. First, the u'v' coordinates of the end point of a vector were determined by

$$u = L_r \cos \theta + u_w \qquad \text{-- (1)}$$

$$v = L_r \sin \theta + u_w \qquad \text{-- (2)}$$

where L_r is the vector length long radii, and θ is the colour direction angle.

The grey reference and vector endpoint were then transposed to x y coordinates using

$$x = \frac{9u}{6u-16v+12} \qquad \text{-- (3)}$$

$$y = \frac{4v}{6u-16v+12} \qquad \text{-- (4)}$$

The elliptical parameters for both the CAD and CCT results were calculated after fitting an ellipse to the data with the centre of ellipse set to the grey background for each test. The elliptical fitting routine was an Octive shareware program (Octive-4.0.0 GUI, downloaded from GNU Octave-GNU.org, Copyright (C) 2015 John W. Eaton and modified by Zach Barnes). There were 12 CVDs results (8 deutans and 4 protans) on CCT test that produced a message from the program stating that the ellipse was not the best conic section to fit the data. The CCT program also provides an elliptical fit to the 8 vectors in u'v' space,

but it does not provide information on the validity of the fit. Because the poor elliptical fit could result either from the change in orientation and spacing of the vectors in x, y space or extreme threshold values in one or multiple directions, an elliptical fit was performed with the original u'v' results for those subjects. The same error message was obtained using the u'v' data. Thus, the poor elliptical fit was likely due to extreme threshold values and so the results from those subjects' were included in the analysis since they could be included in any analyses of the CCT.

5.2.2 Vector length

Because the RCCT, CCT Tri, and LandC were performed monocularly, the thresholds and sensitivities of the right and left eye were compared for each test using a paired t-test. There were no significant differences ($p > 0.15$) between eyes for each cone type and so the results for each subject's right and left eye were averaged.

The thresholds for the 3 directions were converted vectors in x, y space using the following steps. For the CCT Tri test, the u' v' chromaticity coordinates were obtained using the equations (1) and (2) and then was converted to x, y coordinate using equations (3) and (4).

The final step was to calculate the vector length for each cone using the following equations

$$L_{x,y} = \sqrt{(x_w - x_P)^2 + (y_w - y_P)^2} \text{ -------------------------------- (6)}$$

$$M_{x,y} = \sqrt{(x_w - x_D)^2 + (y_w - y_D)^2} \text{ --------------------------- (7)}$$

$$S_{x,y} = \sqrt{(x_w - x_T)^2 + (y_w - y_T)^2} \text{ ------------------------- (8)}$$

87

where $L_{x,y}$, $M_{x,y}$, and $S_{x,y}$ are the vector lengths along the protan, deutan and tritan directions for the CCT Tri test or the vector length along L, M S directions in the x y chromaticity diagram. The x_w and y_w are the chromaticity coordinates of the grey background. The x_P and y_P are the chromaticity coordinates for the protan or L cone threshold. The x_D and y_D are the chromaticity coordinates for the deutan or M cone threshold. The x_T and y_T are the chromaticity coordinates for the tritan or S cone threshold.

The LandC test cone log contrasts were converted to x y coordinates using the following steps. The background luminance and chromaticity coordinates were measured using PR-670 Spectroradiometer (Photo Research, Syracuse, NY). The cone excitation was calculated for the grey background using Stockman and Sharpe matrix given by Golz and Macleod (2003)

$$
\begin{array}{l} Lw \\ Mw \\ Sw \end{array} = \begin{array}{ccc} 0.17156 & 0.52901 & -0.02199 \\ -0.15955 & 0.48553 & 0.04298 \\ 0.01916 & -0.03989 & 1.03993 \end{array} * \begin{array}{l} X \\ Y \\ Z \end{array} \quad \text{------------------------- (9)}
$$

where L_w, M_w, and S_w are the L, M, and S cone excitation respectively for the grey background.

The next step was to calculate the cone excitation for each cone using the subject's threshold and the grey point cone excitation using the following equations

$$L = L_w * (10^\wedge L_{cth}) + L_w \qquad \text{---------------------------------- (10)}$$

$$M = M_w * (10^\wedge M_{cth}) + M_w \qquad \text{------------------------------------ (11)}$$

$$S = S_w * (10^\wedge S_{cth}) + S_w \qquad \text{------------------------------------ (12)}$$

Where L_w, M_w, S_w are the cone excitation for the grey point and L_{cth}, M_{cth}, and S_{cth} are the L, M, and S cone thresholds respectively.

The third step was to calculate the tristimulus values (X, Y, and Z) from the L, M or S excitation using the inverse of the above LMS matrix.

$$\begin{matrix} X \\ Y \\ Z \end{matrix} \quad = \quad \begin{matrix} 2.892 & -3.135 & 0.191 \\ 0.952 & 1.021 & -0.022 \\ -0.017 & 0.097 & 0.957 \end{matrix} \quad * \quad \begin{matrix} L \\ M \\ S \end{matrix} \qquad \text{-------------------------------------- (13)}$$

The chromaticity coordinates were calculated using the following equations

$$x = \frac{X}{X+Y+Z} \qquad \text{--- (14)}$$

$$y = \frac{Y}{X+Y+Z} \qquad \text{--- (15)}$$

$$z = 1-x-y \qquad \text{--- (16)}$$

Finally, equations 6, 7 and 8 were used to calculate the vector length from the chromaticity coordinates.

The RCCT test estimates the relative sensitivity of each cone type. Because the letters were only presented for a short time, 5 seconds, it was not possible to measure the chromaticity coordinates of each letter. Instead, the sensitivities were converted to log RCCT (threshold) using the minimum and maximum contrast level reported by Rabin et al. (Rabin et al., 2011). The cone contrasts levels were 27.5% to 1% for the L and M cone and 173% to 7% for the S cone. Because there were 25 letters for each cone, each letter was assigned a

value of 0.05 log unit so that the threshold was estimated by counting the number of correct or incorrect responses.

To calculate the chromatic threshold, first the number of letters missed is needed. This can be calculated by:

$$Letter\ missed = \frac{100 - A}{5}$$ --- (14)

where 100 is the maximum relative sensitivity and A is the subject score out of 100. The log cone contrast threshold (log RCCT) was calculated for L and M cone by

Log RCCT (threshold) = letter missed x 0.05 – 2.00------------------- (15)

And for S-cone by

Log RCCT (threshold) = letter missed x 0.05 – 1.1-------------------- (16)

The luminance and chromaticity coordinates of grey background were measured using the PR-670 Spectroradiometer. The cone excitation was calculated for the grey background using equation 9. Next, the cone excitation was calculated for each cone using the subject's threshold and the grey point cone excitation with equations 10, 11, and 12. Equation 13 was used to calculate the tristimulus values (X, Y and Z) from the LMS thresholds. Next, the chromaticity choordinate were calculated using equations (14-15). The final step was to calculate the vector length from the chromaticity coordinates using Equations 6 through 8.

5.3 Subjects

The study recruited 60 CVNs and 68 CVDs. Section 3.1 describes the subject demographics.

5.4 Results

5.4.1 Chromatic threshold measurement

5.4.1.1 Elliptical parameters

Figures 31 and 32 show examples of the fitted ellipses along with the data points from the CAD and CCT tests respectively. The examples are a CVN (top), deuteranope (middle), and protanope (bottom). As expected the CVN is smaller than the dichromat ellipses, and the orientation of the ellipses differ with the CVN orientated in parallel to a tritan line of confusion and the red-green dichromat ellipses orientated parallel to their respective lines of confusion through the grey reference.
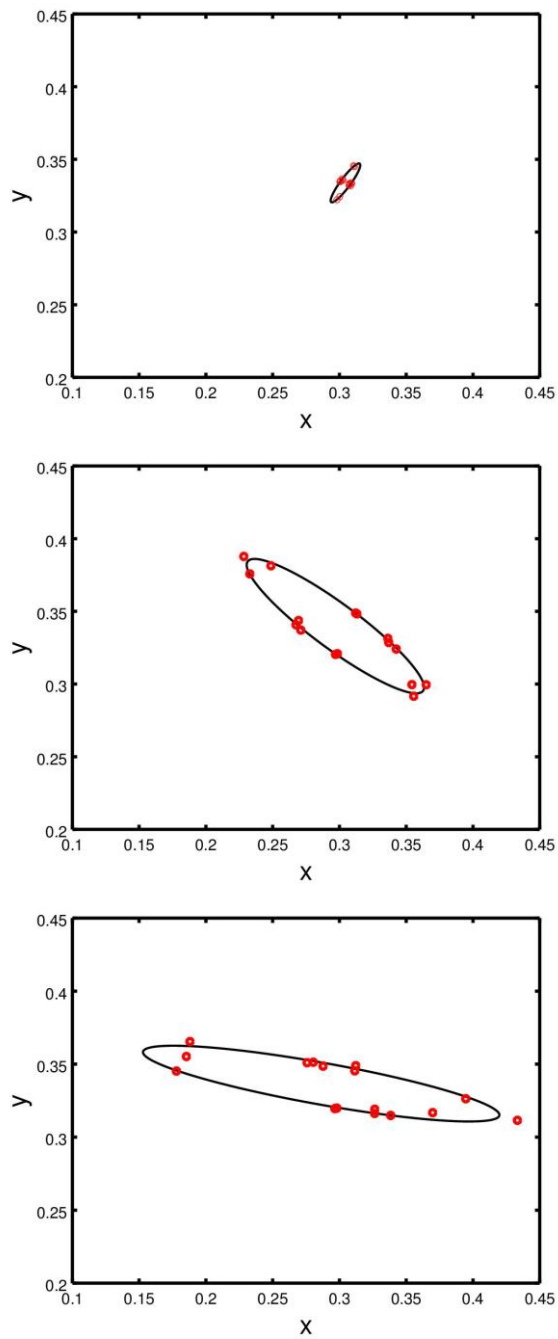
Figure 31. Examples of ellipses of a normal (top), deuteranope (middle) and protanope (bottom) subject from CAD test. The red dots are the data points, and the black locus is the fitted ellipse.
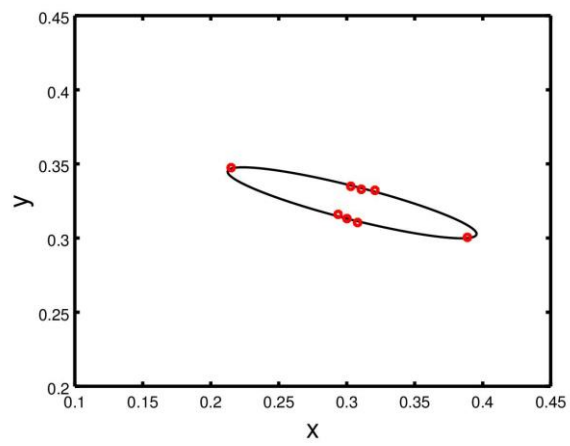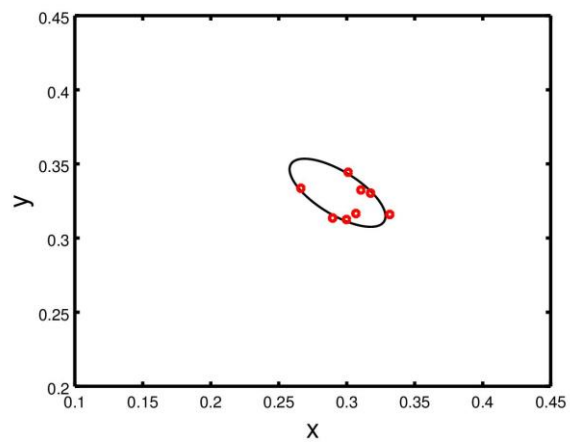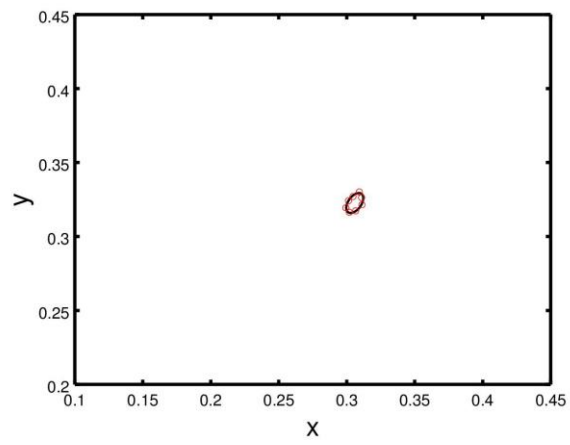
Figure 32. Examples of CCT ellipses of the same normal (top), deuteranope (middle) and protanope (bottom) subjects shown in the previous figure. The red dots are the data points, and the black line is the fitted ellipse.

Figure 33 shows CAD and CCT elliptical area for the CVNs and CVDs. In general, the mean area was largest for the dichromats, and the anomalous trichromatic ellipses were larger than CVNs. There was considerable overlap of the anomalous trichromats with CVNs and their corresponding dichromats. One deuteranope and two protanopes had results on the CCT that were within the CVN range.

Regarding the difference in size measured by the two tests, the average CVN elliptical area ($3.3*10^{-4}$, SD $\pm5*10^{-4}$) on the CAD test was larger than the average area ($1.9*10^{-4}$, SD $\pm2*10^{-4}$) on the CCT test. Similarly, the average CVD elliptical area on CAD test ($5.65*10^{-3}$, SD $\pm7*10^{-3}$) was larger than the CCT test ($3.4*10^{-3}$, SD $\pm4*10^{-3}$). For all statistical analyses in this chapter, the the anomalous trichromats and dichromats data were pooled together into protan and deutan groups because of the small number of dichromats and the overlap of data. Repeated measures analysis of variance for the CVNs, protans, and deutans showed that there was a significant effect of test ($F_{(1)} = 5.88$; $p=0.017$) and colour vision type ($F_{(2)} = 26.735$; $p<0.05$). However, the interaction term was not significant ($F_{(2)} = 1.462$; $p=0.236$). This indicates that the three groups were performing the two tests in the same manner, with the protans having areas larger than the deutans and CVN. The Greenhouse-Geisser adjustment was used for the within-subjects effects and interaction because the data did not meet the sphericity assumption.
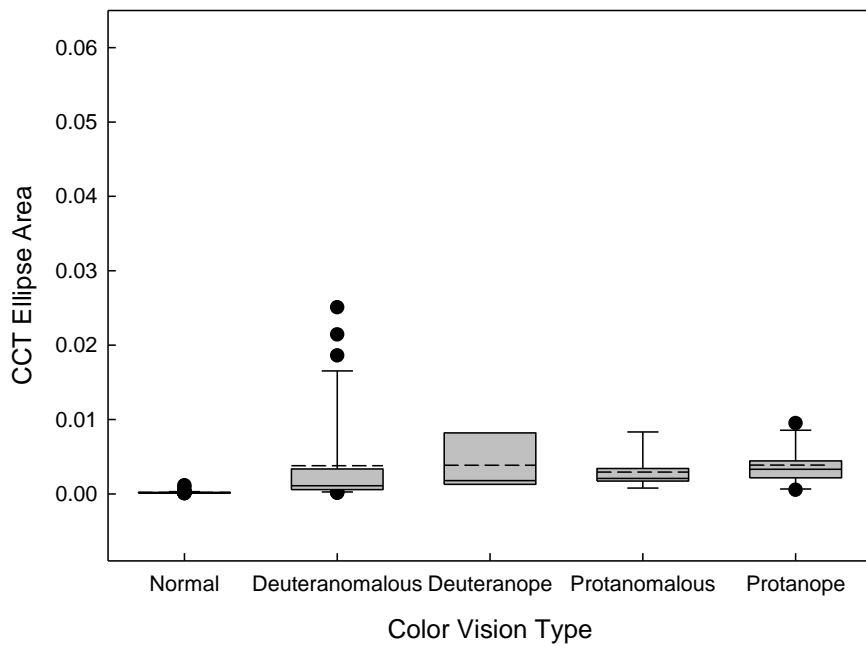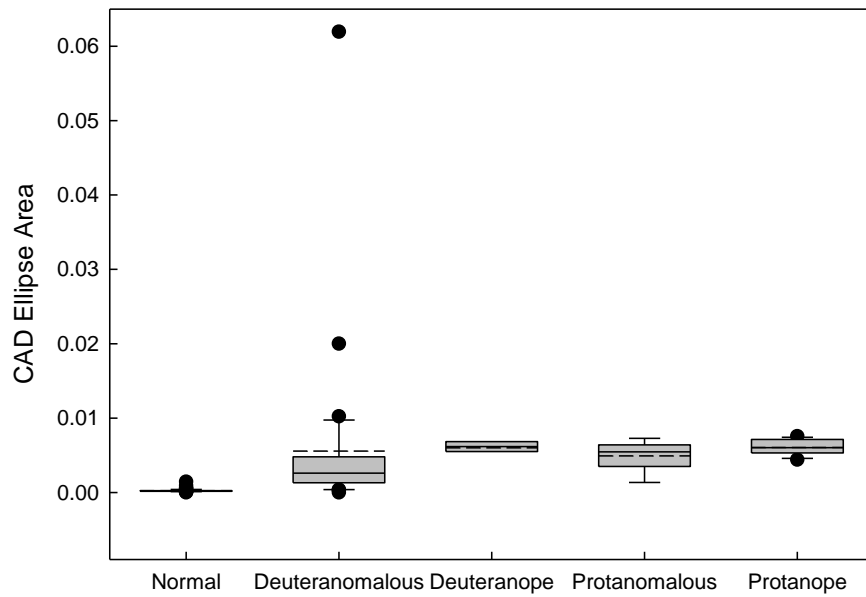
Figure 33. The ellipse area results for CVN and CVD subjects for the CAD (top) and CCT (bottom) test. The dash line represents the mean and solid lines are the median value. The box limits are the 25th and 75th percentiles and the error bars are 10th and 90th percentiles. The black dots are the outlier results.

Figure 34 shows CAD (top) and CCT (bottom) elliptical angle results for the CVNs and CVDs. For the angle analysis, 180 degrees was added to all angles less than 90 degrees so that all the angles fell within the second and third quadrants. Using this transformation, the angles for the CVDs were smaller than the CVNs. However, the key difference was that the CVN were generally orientated towards the vertical axis and CVD were oriented near horizontal axis as expected. With one exception, the deuteranopic angles did not overlap with the protanope and protanomalous data on either test. The exception was one protanope on the CCT test who had an angle within the deuteranopic range. However, the deuteranomalous angles were variable and overlapped with the other groups on both tests.

Repeated measures analysis of variance with the Greenhouse-Geisser correction for sphericity showed that the effect of tests ($F_{(1)} = 2.501$; p=0.116) and interaction ($F_{(2)} = 0.011$; p=0.989) were not significant. However, the effect of the type of colour vision was significant ($F_{(2)} = 291.119$; p<0.05). The lack of a significant interaction indicates that the three groups were performing the two tests in the same manner with the CVN having the larger angle followed by the protans and the deutans' mean angle was the smallest.
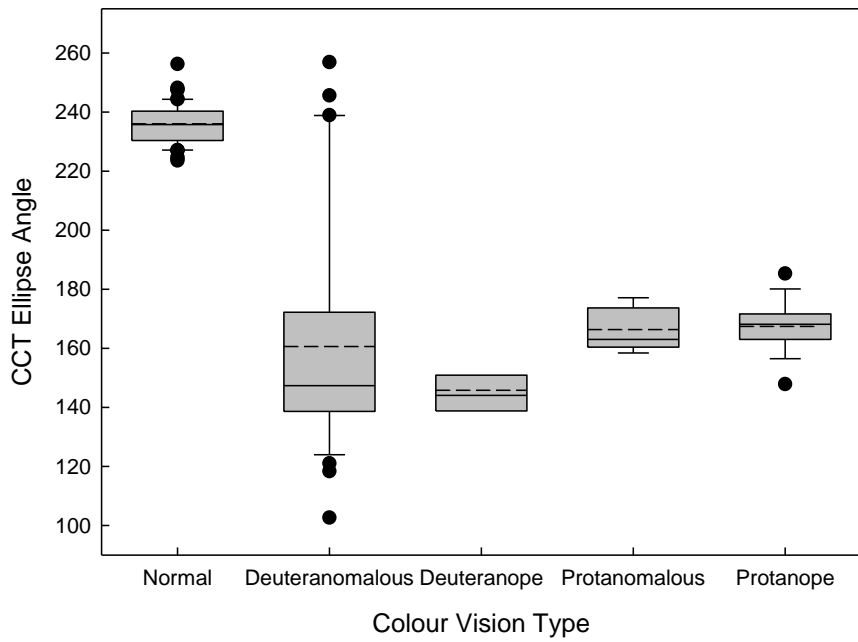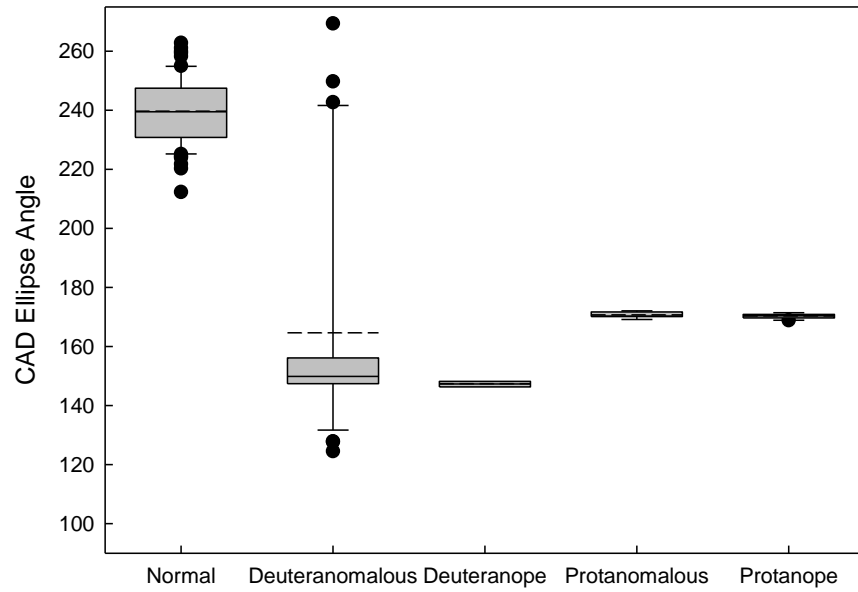
Figure 34. The ellipse angle results for CVNs and CVD subjects for the CAD (top) and CCT (bottom) test. The dash line represents the mean and solid lines are the median values. The box limits are the 25th and 75th percentiles and the error bars are 10th and 90th percentiles. The black dots are the outlier results.

5.4.1.2 Vector Length measurement

Figure 35 shows the L cone (protan) vector length results for each colour vision test. As expected, the L-cone (protan) thresholds for the protan were the highest across all tests. Only one protanope had an L-cone threshold within the normal range, and that was on the RCCT. This subject passed the RCCT test with his right eye and failed it with his left eye. The somewhat surprising result was that the deutan results for the CCT showed an elevated threshold along the protan direction. Although the mean values were lower than the protan values, the deutan results did overlap with the protan thresholds.

For statistical analysis, the CVNs were analyzed separately from the CVDs because there is a floor effect for the CVN with the lowest red-green contrast on the RCCT above the threshold for many colour-normals as evident by the LandC threshold.

In the CVN group, the median L-cone vector length was the largest on the CCT Tri followed by the RCCT, and the LandC was the smallest. The differences between the tests were statistically significant as determined by Friedman test (X^2= 83.733; df= 2; p<0.005). Both the protans and deutans had largest average L-cone vector length on the CCT Tri followed by the LandC and then the RCCT. Repeated measures analysis of variance showed that the effect of tests ($F_{(1.425)}$ = 107.3; p<0.005) and type of defect were significant ($F_{(1)}$ = 445.997; p<0.005). The interaction term was also significant ($F_{(1.425)}$ = 35.96; p<0.005). The

significant interaction term was due to deutans having highter L-cone thresholds on CCT Tri

relative to the L-cone thresholds on the other two tests, which were similar to the CVNs's

results. The Greenhouse-Geisser adjustment was used for the within-subjects effects and

interaction because the data did not meet the sphericity assumption.
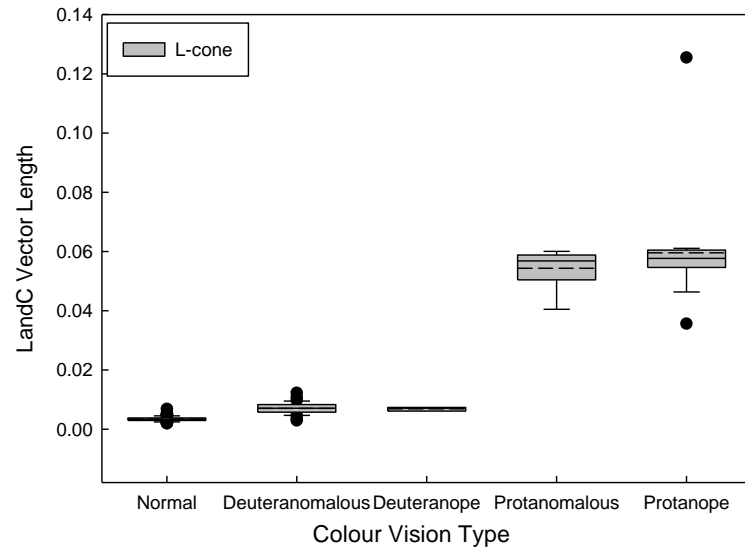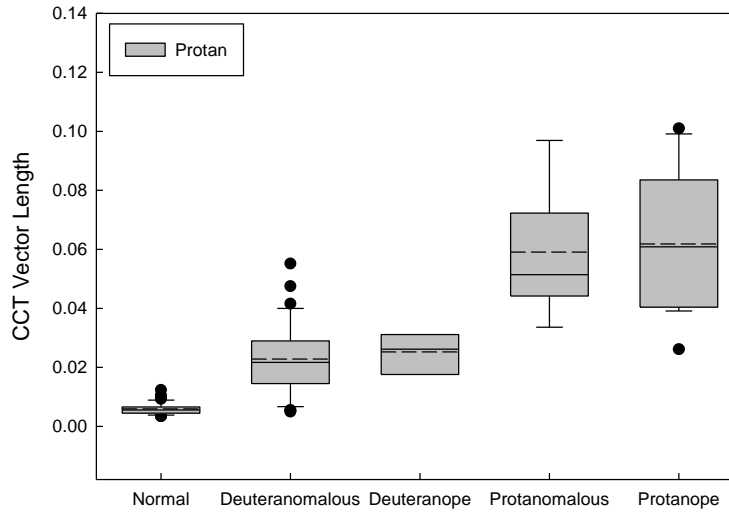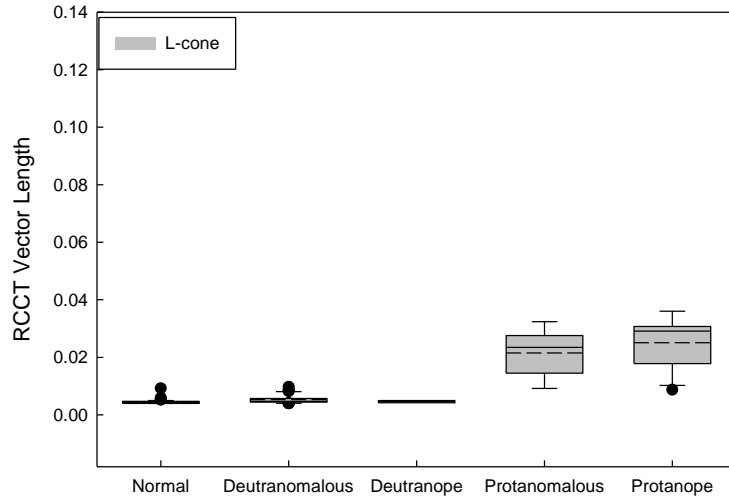
Figure 35. The L cone (protan) vector length results for the RCCT (top), CCT Tri (middle) and LandC (bottom) test for CVNs and CVD. The dash line represents the mean and solid lines are the median values. The box limits are the 25th and 75th percentiles and the error bars are 10th and 90th percentiles. The black dots are the outlier results.

Figure 36 shows the M cone (deutan) vector length results as a function of each colour vision test. The M-cone results for the LandC and RCCT showed that the deutan individuals had the higher thresholds, whereas the protans had thresholds within the normal range. This indicates that the tests had a strong ability to separate the protans from deutans. The exceptions were three deuteranomalous subjects who had a vector length within the normal range on both tests. All the three deuteranomalous passed both tests using each eye. For the CCT Tri, however, the thresholds for the protans along the deutan line of confusion were higher than the thresholds for the deutan subjects, In fact, the protan mean values for deutan vector were higher than their thresholds along the protan direction. None of the protans had a deutan threshold value within the normal range.

In terms of the differences between tests in the CVN group, the largest median M cone vector length was on the CCT Tri followed by the RCCT and then the LandC. The differences were a statistically significant between tests as determined by Friedman test ($X^2 = 75.233$; df= 2; p<0.005). For the CVD groups, the differences between tests for the M-Cone were similar to the L-cone results with the largest average M-cone vector length on the CCT Tri and slightly smaller on the LandC and then the lowest on the RCCT. In terms of the differences within tests, deutan subjects had an average M-cone vector length larger than the

protans on the RCCT and LandC, whereas the protan individuals' average M cone vector length was larger than the deutans on the CCT Tri. Repeated measures analysis of variance for the CVDs showed that the effect of tests ($F_{(1.159)} = 312.774$; $p<0.005$) was significant. However, the effect of the type of defect was not significant ($F_{(1)} = 3.482$; $p<0.066$). The interaction term was significant ($F_{(1.159)} = 148.826$; $p<0.005$) due to the higher M-cone thresholds for the protan subjects on the CCT Tri and normal M-cone thresholds on the other tests. The Greenhouse-Geisser adjustment was used for the within-subjects effects and interaction because the data did not meet the sphericity assumption.
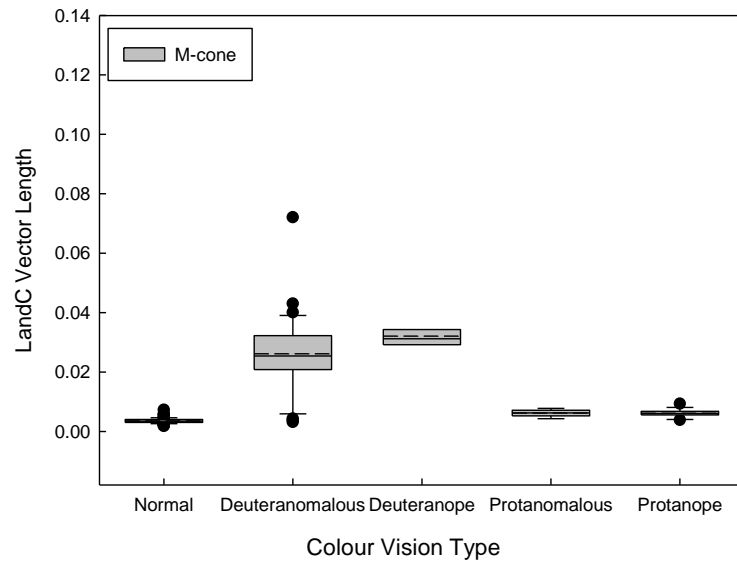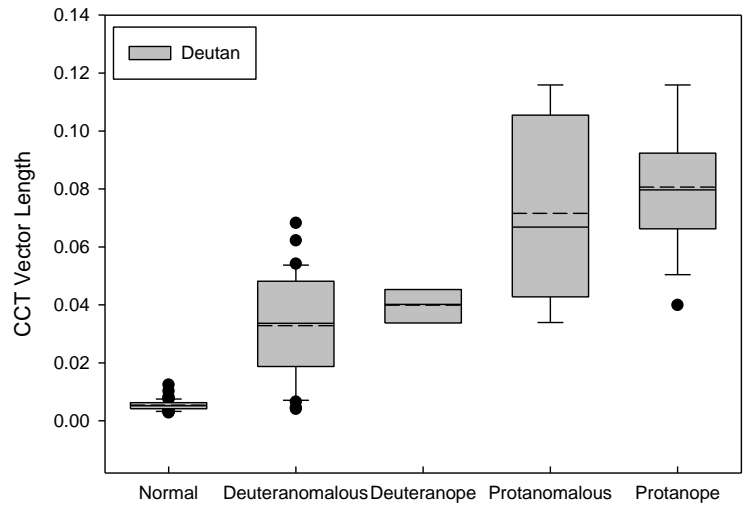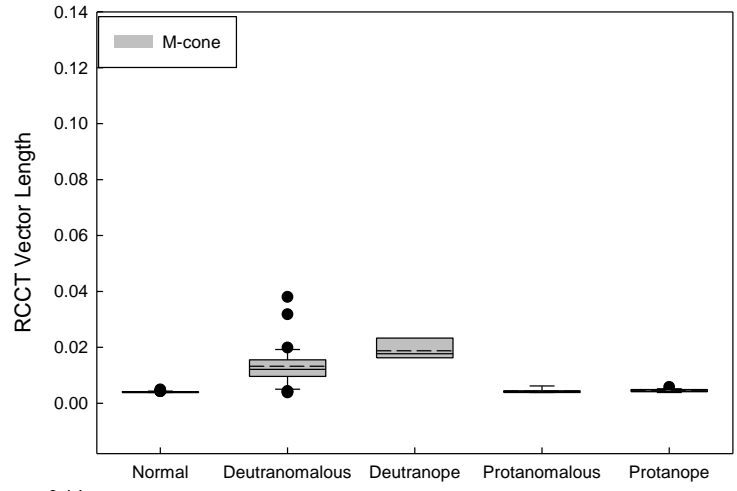
Figure 36. The M cone (deutan) vector length results on the RCCT (top), CCT Tri (middle) and LandC (bottom) test for CVNs and CVD. The dash line represents the mean and solid lines are the median values. The box limits are the 25th and 75th percentiles and the error bars are 10th and 90th percentiles. The black dots are the outlier results.

Figure 37 shows the S-cone (tritan) vector length results. Within each test, the median was generally the same and ranges overlapped with each other. These results were expected because there were no tritan individuals in the study. Nevertheless, there were some exceptions. One protanomalous and 3 deuteranomalous subjects had a higher threshold than the normals limit on the CCT Tri tritan vector. One of the 3 deuteranomalous subjects also had a higher threshold than the normals on the LandC S-cone vector.

The results between tests were similar across the three groups of subjects. The largest median S-cone vector length was on the LandC, followed by the CCT Tri with the RCCT having the lowest median value. The Friedman test was used to determine whether there were significant differences between results since there is a floor effect for all subjects. The differences between tests were significant for CVN (X^2= 32.033; df= 2; p<0.005), deutans (X^2= 40.55; df= 2; p<0.005) and protans (X^2= 15.5; df= 2; p<0.005).
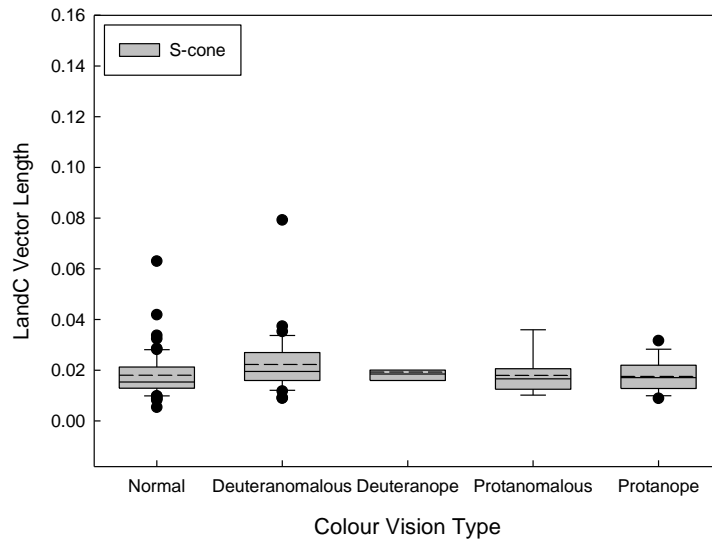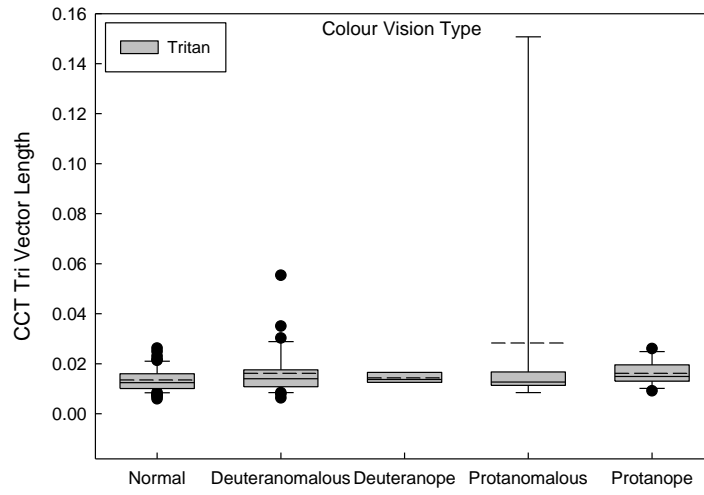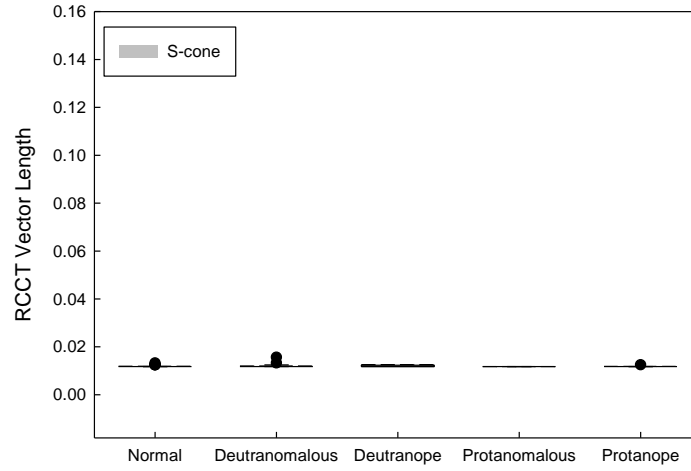
Figure 37. The S cone (tritan) vector length results on the RCCT (top), CCT Tri (middle) and LandC (bottom) test for CVNs and CVDs. The dash line represents the mean and solid lines are the median values. The box limits are the 25th and 75th percentiles and the error bars are 10th and 90th percentiles. The black dots are the outlier results.

5.4.2 Correlation between Tests

5.4.2.1 Elliptical parameters

Although the average threshold values differed between tests, it is worthwhile to know whether the individual threshold values correlate across tests. If the correlation is high, then multiple tests could be used for a given industry rather than just one specific test. Figure 38 and 39 show scatter plots for the CAD and CCT ellipse area and angle results for the two subject groups along with the results of the Spearman Rank correlations. The Spearman Rank was selected over the Pearson correlation because the data were not normally distributed. There was generally moderate and statistically significant correlation between the two tests in terms of the area for both the CVN and CVD groups; however, the angle correlation was only significant for the CVD group. This result was expected because the range in the angles for the CVN was relatively small. The CVN angle ranges were 212º to 262º for CAD and 223º to 256º for CCT.
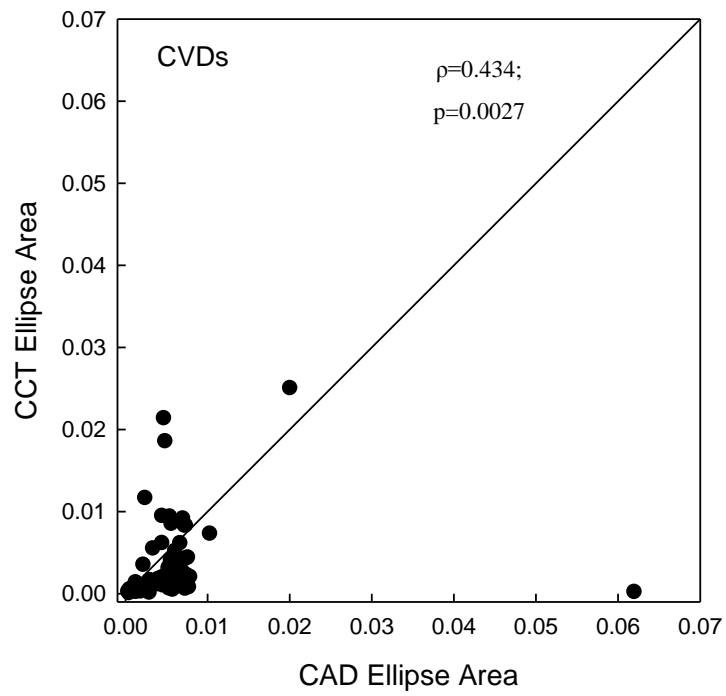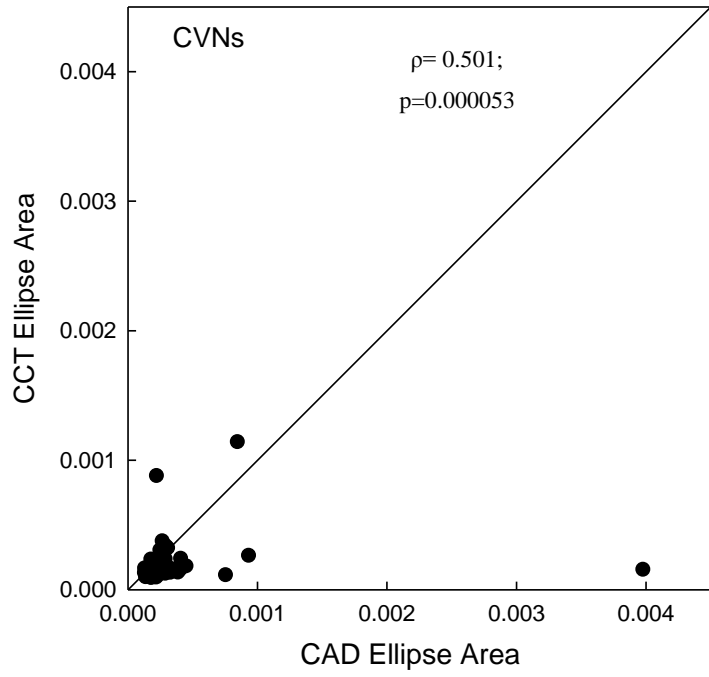
Figure 38. Scatter plot for the CAD and CCT ellipse area results for CVNs (top) and CVDs (bottom). The solid line is the one to one correlation line, ρ is the Spearman Rank value and p is the p-value for the correlation.
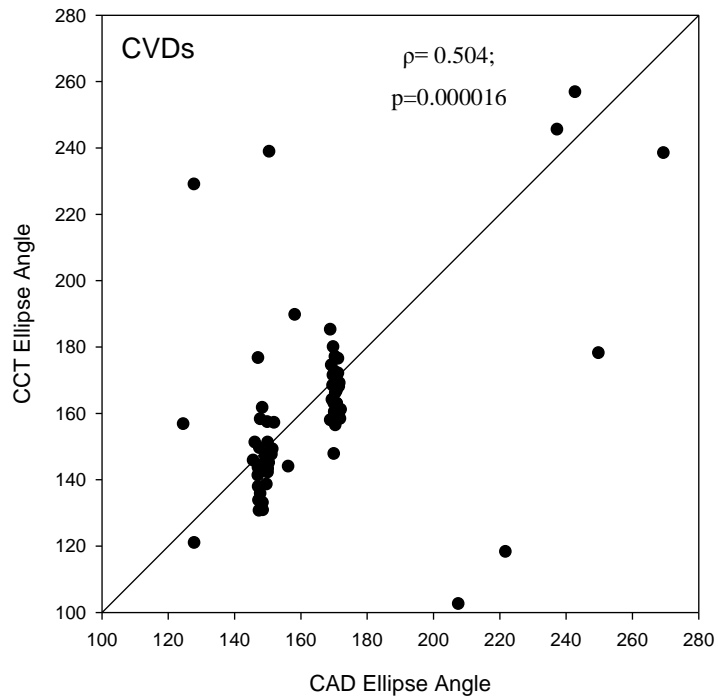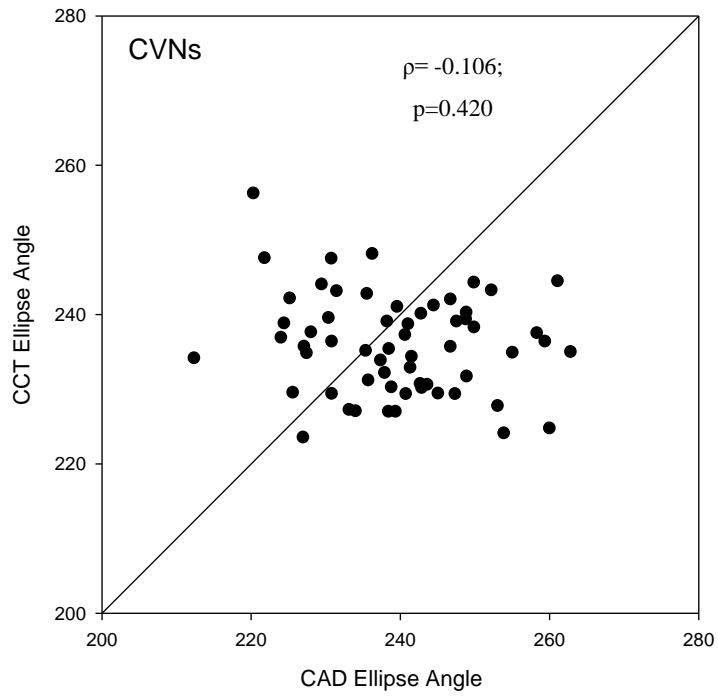
Figure 39. Scatter plot for the CAD and CCT ellipse angle results for CVNs (top) and CVDs (bottom). The solid line is the one to one correlation line, ρ is the Spearman Rank value and p is the p-value of the correlation.

5.4.2.2 Vector Length measurement

Table 3 lists all of the possible Spearman Rank correlation results between the three vector length tests. The CVN group did not have many significant correlations, and those that were significant were only moderate. This was expected based on the limited range of threshold values. For the CVD, the significant correlations were strongest for the M-cone and L-cone thresholds. Nevertheless, there were two exceptions. These were the between the CCT Tri deutan thresholds and the RCCT and LandC M-cone thresholds. The correlations were only moderate, but more importantly, the correlations coefficient were negative.

Table 3. The vector length correlation results between the LandC, RCCT and CCT Tri tests for CVNs and defectives.

		Colour Normals	**Colour Defectives**
LandC	L cone vs L cone	ρ =0.396; p=0.002	ρ =0.844; p<0.001
vs	M cone vs M cone	ρ =0.203; p=0.119	ρ =0.884; p<0.001
RCCT	S cone vs S cone	ρ =0.303; p=0.019	ρ = 0.138; p=0.260
LandC	L cone vs Protan	ρ =0.145; p=0.267	ρ =0.776; p<0.001
vs	M cone vs Deutan	ρ =0.303; p=0.019	ρ = -0.379; p=0.0015
CCT Tri	S cone vs Tritan	ρ =0.444; p=0.0004	ρ =0.178; p=0.146
RCCT	L cone vs Protan	ρ =0.246; p=0.058	ρ =0.693; p<0.001
Vs	M cone vs Deutan	ρ =0.172; p=0.187	ρ = -0.456; p<0.001
CCT Tri	S cone vs Tritan	ρ =0.178; p=0.173	ρ =0.218; p=0.074

Figure 40 shows an example of a scatter plot of the L- cone LandC vector length results as a function of the RCCT L-cone vector length results for CVNs (top) and CVDs (bottom). This is an example of data that had statistically significant correlations for both the CVN and CVD groups. On the CVN, there was a floor effect on the RCCT with no values below $3.99*10^{-3}$. This was the reason for only a fair correlation. The CVDs shows two distinct groups. The data cluster at the lower left corner is the deutan results who had near normal L-cone results and the protan results are the scattered more widely on the right side of the plot. The two distinct threshold areas are responsible for the strong correlation between the two tests.
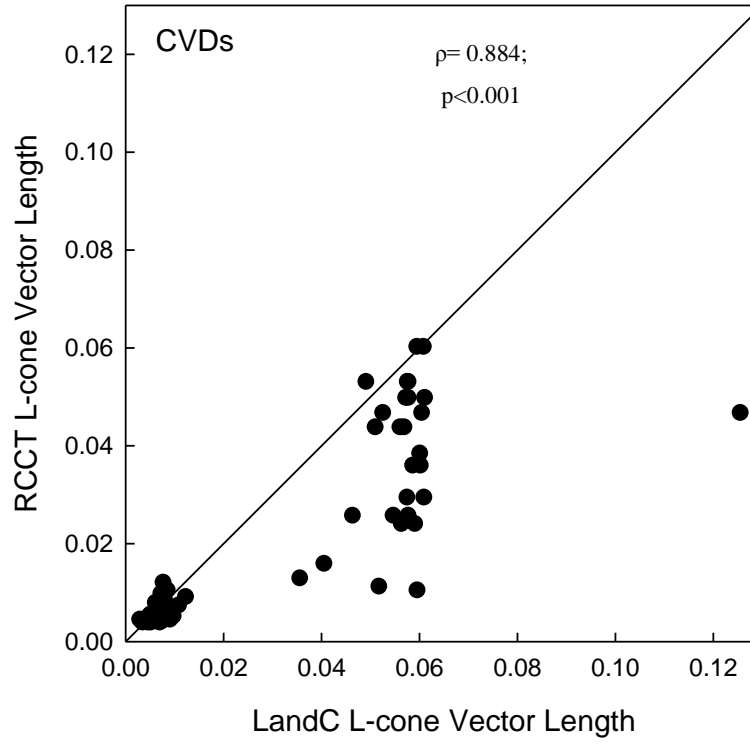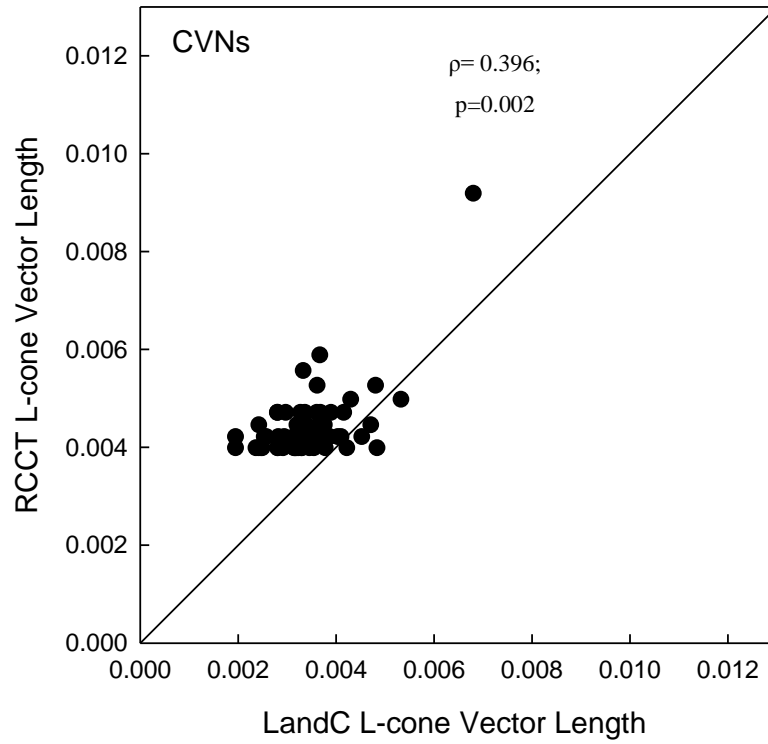
Figure 40. Scatter plot for the LandC L-cone and RCCT L-cone vector length results for CVNs (top) and CVDs (bottom). The sloid line is the one to one correlation line and ρ is the Spearman Rank value along with p-value.

Figure 41 shows a scatter plot as an example of the LandC M-cone and CCT Tri deutan vector length results for CVN (top) and CVD (bottom). This example illustrates a significant positive correlation for the CVN group and a negative correlation for the CVD group. The CVN correlation was moderate and the higher CVN CCT Tri values are expected based on the previous analysis. The negative correlation for the CVDs was primarily a result of the protans who had relatively normal LandC M cone contrast thresholds, but high thresholds on the CCT Tri deutan vector.
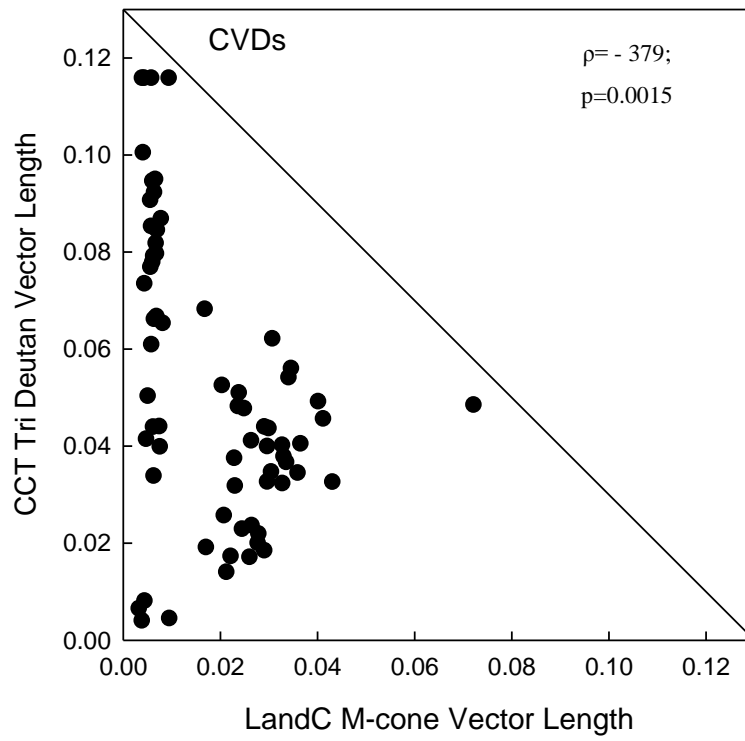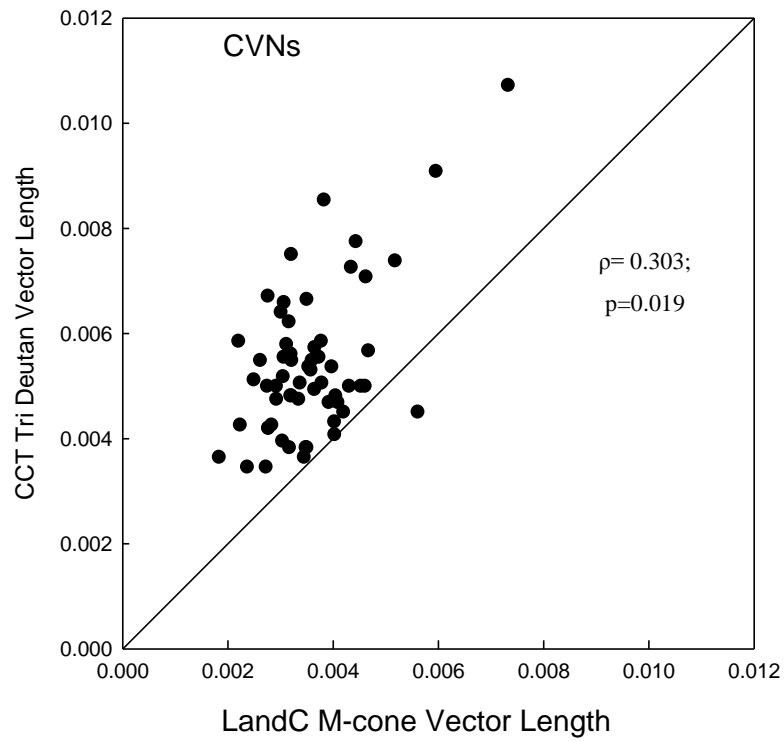
Figure 41. Scatter plot for the LandC M-cone and CCT Tri deutan vector length results for CVNs (top) and CVDs (bottom). The solid line is the one to one correlation line and ρ is the Spearman Rank value along with p-value.

Figure 42 illustrates the case where there is no significant correlation between tests for both the CVN and CVD group. The RCCT S-cone vector length results were presented as a function of the CCT Tri tritan vector length for CVNs (top) and CVDs (bottom). There was a clear floor effect on the RCCT CVN and CVD results, which is the reason for the lack of any association between the RCCT S-cone thresholds and the other two tests.
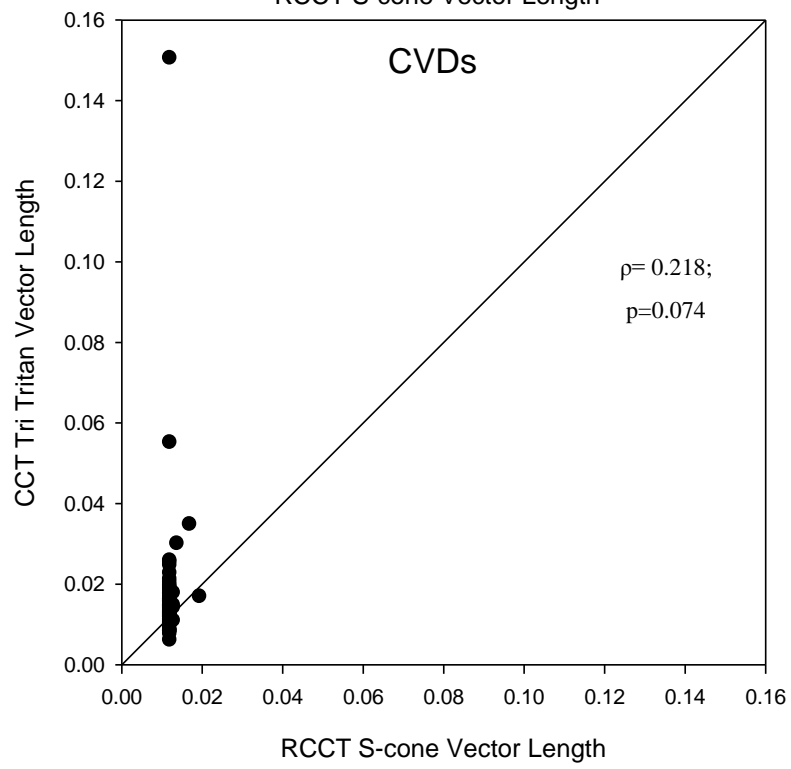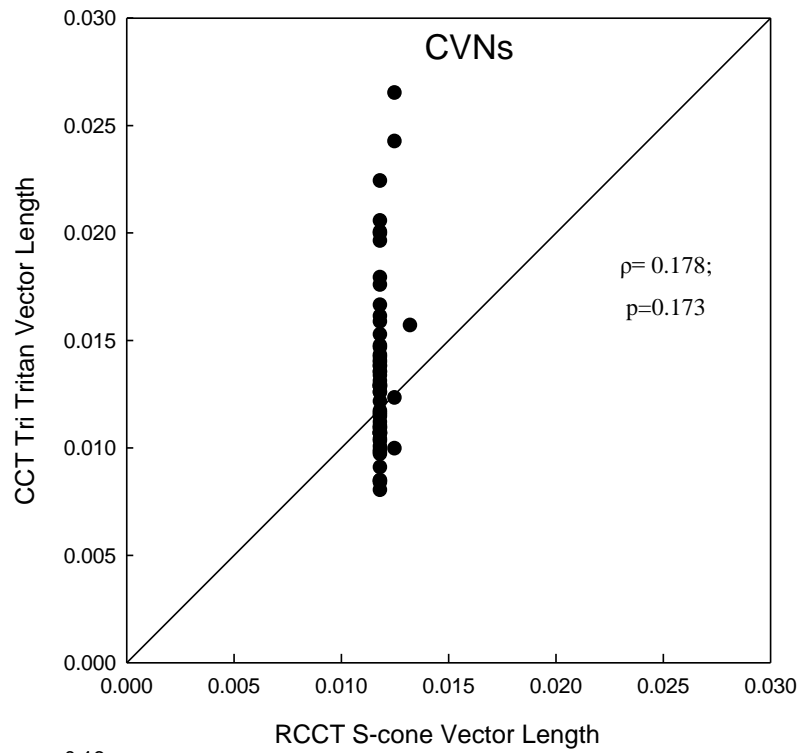
Figure 42. Scatter plot for the RCCT S-cone and CCT Tri Tritan vector length results for CVNs (top) and CVDs (bottom). The solid line is the one to one correlation line and ρ is the Spearman Rank value along with p-value.

5.4.3 Correlation with the Anomaloscope

5.4.3.1 Elliptical parameters

One of the interests on this chapter was to determine whether the ellipse area would correlate with the Rayleigh range of matches. Both parameters are measures of chromatic discrimination. Table 4 shows the Spearman Rank Correlation for the CAD and CCT (ellipse area) with the Rayleigh matching range. There was no significant correlation for CVN group. The correlations were significant for the CVD group, but correlation with the CAD was only moderate and fair for the CCT.

Table 4. Spearman correlation results between the Rayleigh matching range and the elliptical area for CAD and CCT for CVNs and defectives.

	CAD		CCT	
	ρ	P value	ρ	P value
Colour Normal	0.179	0.17	0.046	0.73
Colour Defective	0.495	0.00002	0.254	0.037

The other factor that we were interested in is whether the ellipse area and angle of the CAD and CCT would agree with Rayleigh result (range and mid-point) as to whether the

117

person had a red-green CVD or normal colour vision. Because the cut-off value (area or angle) that would separate the CVNs from the CVDs were not established, Receiver-Operator-Characteristic (ROC) analyses (Sigmaplot ver 11.0; Systat Software Inc, Chicago, IL) were performed to determine the optimum value that could separate the CVNs from the defectives for each test. The optimum cut-off score was based on the maximum sum of the sensitivity and specificity. If there were multiple criteria with the same maximum value, then the cut-off score was selected based on which one of the choices had the highest sensitivity.

The area under the ROC curve for the CAD area was 0.97 (95% CI = 0.94 to 1.00) and 0.96 (95% CI = 0.91 to 1.00) for the CAD angle. The optimum cut-off points were $9.75*10^{-4}$ for CAD area and 209.9º for the CAD angle. The AC1 level of agreement with the Rayleigh classification (or diagnosis) was calculated. The AC1 values can vary from -1 to 1, with -1 indicating complete disagreement, 0 meaning that any agreement is due to chance and 1 indicates perfect agreement. The agreement values were calculated using Agree Stat version 2013.2 (Advanced Analytics, Gaithersburg, MD, USA) (Gwet, 2008; Wongpakaran et al., 2013). Table 5 shows the pass/fail agreement between the Rayleigh match and CAD area and angle. The AC1 coefficient agreement was 0.94 (95% CI= 0.88 to1) for the area and 0.92 (95% CI= 0.85 to 0.99) for the angle. Although the CAD ellipse area agreement was slightly higher than the CAD angle, the AC1 values were not significantly different based on the 95% confidence interval. If we use either the area or the angle, the AC1 agreement with the Rayleigh match was 0.91 (95% CI= 0.83 to 0.98). The discrepancies were all the 5

subjects who failed the Rayleigh match but passed the CAD angle plus an additional subject who passed the Rayleigh but failed CAD using the elliptical area.

Table 5. The agreement between Rayleigh match and the two CAD elliptical parameters as whether the person has a normal or colour vision defect.

		Rayleigh	
		Normal	Red-Green Defective
CAD area	Normal	59	3
	Red-Green Defective	1	64
CAD angle	Normal	60	5
	Red-Green Defective	0	62

For the CCT test, the area under the ROC curve for the ellipse area was 0.97 (95% CI = 0.94 to 1.00) and 0.96 (95% CI = 0.91 to 1.00) for the angle. The optimum cut-off point was $4.54*10^{-4}$ for CCT area and 206.7° for CCT angle. Table 6 shows the pass/fail agreement between the Rayleigh results and CCT area and angle. The agreements coefficient with the Rayleigh classification were good with an AC1 value of 0.86 (95% CI= 0.77 to 0.95) for the ellipse area and 0.92 (95% CI= 0.85 to 0.99) for the ellipse angle. Although the AC1 agreement was lower for the ellipse area, it was not significantly different from the angle AC1 based on the 95% confidence interval. If we consider using either the ellipse area or angle, the agreement with Rayleigh match was exactly the same as the AC1 value calculated

for elliptical area. This was because all the discrepancies when using the ellipse area included all the individuals who had discrepancies using the elliptical angle.

The AC1 agreement values of the Rayleigh match with the CAD elliptical angle or CCT elliptical angle were identical. However, the AC1 agreement of the CCT area was slightly smaller than the agreement of CAD area but was not statistically significant based on the 95% confidence interval.

Table 6. The agreement between Rayleigh match and the two CCT elliptical parameters as whether the person has a normal or colour vision defect

		Rayleigh	
		Normal	Red-Green Defective
CCT area	Normal	58	7
	Red-Green Defective	2	61
CCT angle	Normal	60	5
	Red-Green Defective	0	63

Another issue is whether the elliptical angle would agree with Rayleigh diagnosis (based on mid-point and range) as to whether the CVD has a protan or deutan defect. ROC analyses were also performed to determine the optimum cut-off point that could separate the protans from deutan for each test. Only the subjects who failed each test by either ellipse area or angle were included in this analysis. The area under the curve for the CCT was 0.90 (95% CI = 0.80 to 0.99) and 0.92 (95% CI = 0.83 to 1.00) for the CAD test. The cut-off points

were 157.8º for the CCT and 163.5º for the CAD. Table 7 shows the classifications

agreement between the Rayleigh with the CAD and CCT diagnosis results. The AC1

coefficient agreement was 0.97 (95% CI= 0.90 to 1) for the CAD and 0.78 (95% CI= 0.62 to

0.94) with CCT. The AC1 values were significantly different based on the 95% confidence

interval.

Table 7. Comparison of the CAD and CCT classification with the Rayleigh classification.

		Rayleigh	
		Protan	Deutan
CAD	Protan	28	1
	Deutan	0	33
CCT	Protan	26	5
	Deutan	2	30

5.4.3.2 Vector Length

Table 8 shows the Spearman Rank correlations between the Rayleigh matching range and the

RCCT, LandC, and CCT Tri tests for the CVNs and CVDs. Note that the CVNs' Rayleigh

match range results were compared with each cone type (L and M cone) on each test. The

Rayleigh diagnosis (range and mid-point) determined the classification the CVDs as either

deutan or protan; however, the correlations were carried out as follows:

- The deutans matching range was compared with the M cone vector length on RCCT

    and LandC and with the deutan vector length on CCT Tri. The CVD subject who had

a normal Rayleigh match (range and mid-point) and failed all the other tests with a

deutan defect was included in the deutan group.

- The protan matching range was compared with the L cone vector length on RCCT

and LandC and to the protan vector length on CCT Tri.

The correlations for the CVNs were not significant. Only the deutan subjects had a

significant correlation between tests. Their M-cone threshold on the RCCT and LandC tests

had a significant, but moderate correlation, with Rayleigh range. The correlations with the L-

cone thresholds were not significant as were either correlations between the CCT Tri and the

matching range.

**Table 8. Spearman correlation results between the Rayleigh matching range and the vector
length measurements of RCCT, LandC, CCT Tri tests for CVNs and defectives.**

		Colour Normals	**Colour Defectives***
Rayleigh vs RCCT	L cone	ρ =0.105; p=0.425	ρ =0.302; p=0.118
	M cone	ρ =0.134; p=0.305	ρ =0.484; p=0.0017
Rayleigh vs LandC	L cone	ρ =0.141; p=0.28	ρ =0.264; p=0.173
	M cone	ρ =0.103; p=0.433	ρ = 0.433; p=0.005
Rayleigh vs CCT Tri	Protan	ρ =0.074; p=0.573	ρ =0.098; p=0.617
	Deutan	ρ =0.0013; p=0.992	ρ = 0.218; p=0.175

* The deutan matching range was compared with the M cone vector length on RCCT and
LandC and with the deutan vector length on CCT Tri. The protan matching range was
compared with the L cone vector length on RCCT and LandC and to the protan vector
length on CCT Tri.

5.4.3.3 S-cone (Tritan)

Subjects with normal and defective colour vision found that the colour matching on Moreland anomaloscope was a very difficult. The reason for the difficulty was probably that the reference field appeared more saturated than the mixture of the two primaries so that subjects could only obtain a hue and brightness match, but not a saturation match. The midpoint was 53.9 (SD= ±5.9) for the CVNs. However, the CVD group required more green to make a match than the CVN group with a midpoint match of 60.8 (SD=±10.2) and 63.8 (SD=±9.5) for deutan and protan respectively. Figure 43 shows a boxplot for the Moreland midpoint of the range results for colour-normals and defectives (protan and deutan).
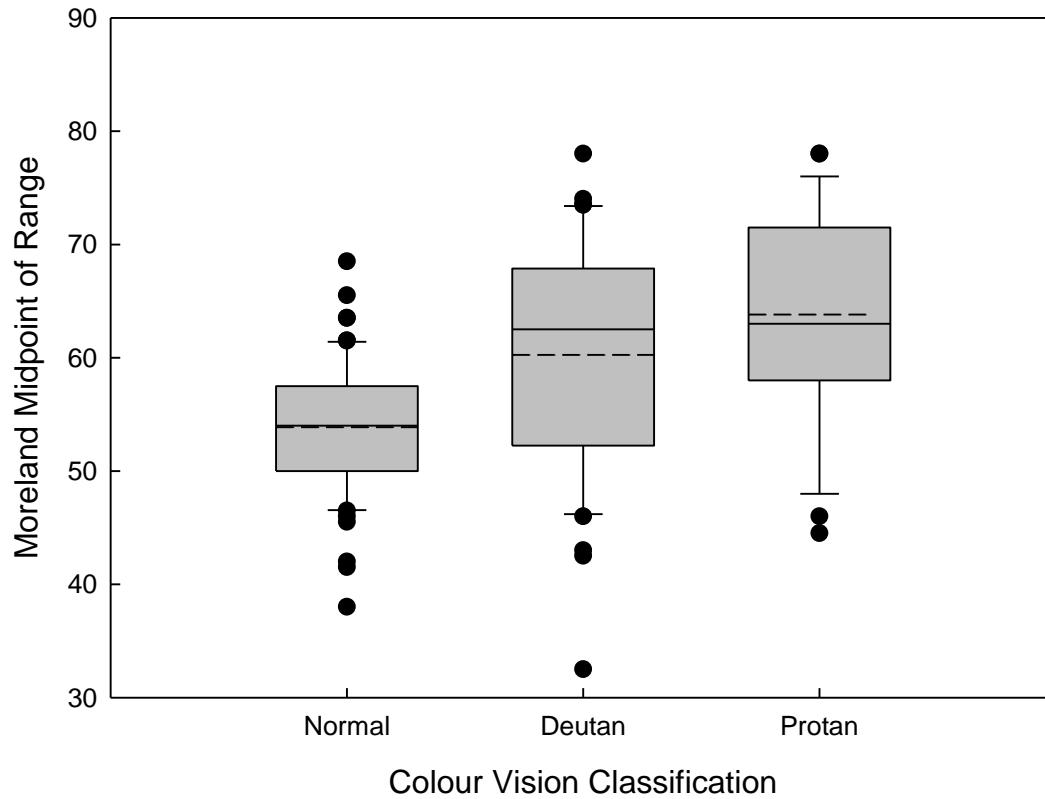
Figure 43. The Moreland midpoint of the range results for CVNs and defectives. The dashed line represents the mean and solid lines are the median values. The box limits are the 25th and 75th percentiles and the error bars are 10th and 90th percentiles. The black dots are the outlier results.

Table 9 shows the Spearman Rank correlation between the Moreland anomaloscope range with the S-cone/tritan vector length measurements for RCCT, LandC, and CCT Tri test. There was no significant correlation between any of the tests. Figure 44 shows scatter plots of Moreland range with the LandC S-cone thresholds for both the CVN and CVD

groups. The data scattered widely and showed no clear correlation pattern. The pattern was similar for the RCCT and CCT Tri tests.

Table 9. Spearman Rank correlation results between the Moreland anomaloscope range with S-cone/tritan vector length measurements of RCCT, LandC, CCT Tri tests for CVNs and defectives.

	Colour Normals	**Colour Defectives**
Moreland vs RCCT	$\rho = 0.206$; $p = 0.114$	$\rho = 0.058$; $p = 0.638$
Moreland vs LandC	$\rho = 0.171$; $p = 0.161$	$\rho = 0.021$; $p = 0.863$
Moreland vs CCT Tri	$\rho = 0.113$; $p = 0.390$	$\rho = -0.021$; $p = 0.861$
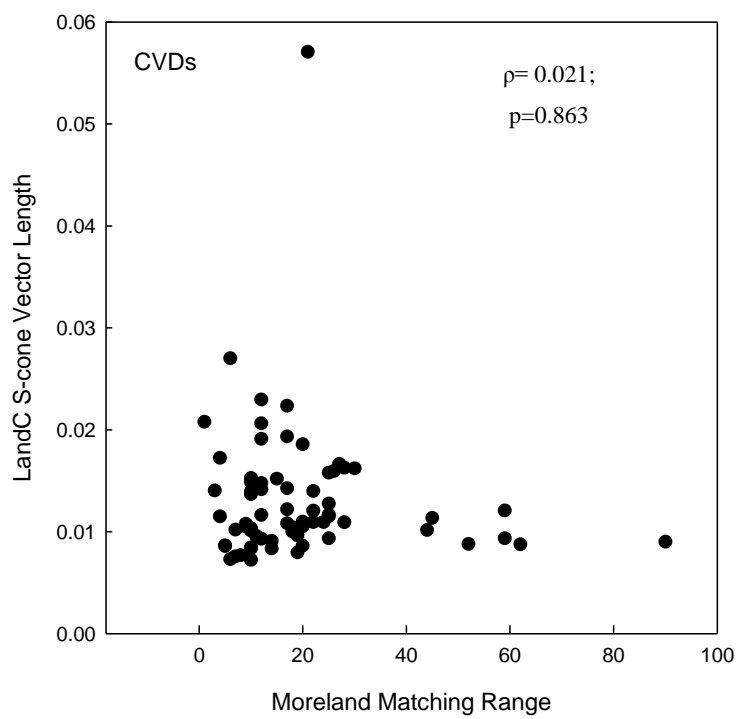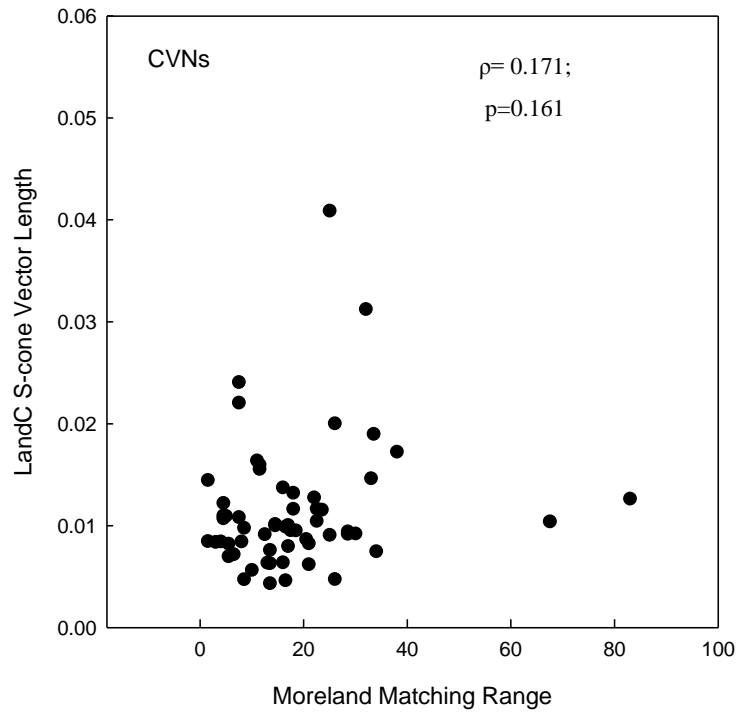
Figure 44. Scatter plot for the LandC S-cone and Moreland matching range results fora CVNs (top) and CVDs (bottom). The ρ is the Spearman Rank value.

5.5 Discussion

Both the CAD and CCT were capable of measuring chromatic discrimination ellipses in a clinical setting. The CAD ellipses were significantly larger than the CCT for both the CVN and CVD groups. The difference was not due to the angular size of the target as both stimuli have about the same angular size of detail. There are a number of other differences between the two tests that could be contributing to the size difference. The CAD test requires identifying motion direction in dynamic noise, whereas the CCT is identification of gap location in static noise. Identifying the motion direction would be expected to have lower threshold (Barbur, 2004). It is also unlikely that the difference was due to differences in the luminance noise used in each test; at least for CVNs. Chromatic thresholds would be similar for CVN subjects whether the noise is static or dynamic (Linhares et al., 2016; Birch et al 1991). The influence of the two types of luminance noise on CVD thresholds is less certain. One study found that dichromats had a larger threshold with the dynamic noise relative the static noise (Birch et al., 1991) and another study found that the anomalous trichromatic had a lower threshold when dynamic noise was used relative to the static noise (Linhares et al., 2016). One possible explanation for the difference in CVD results was that the size and temporal frequency for the dynamic luminance contrast noise used in these two studies were different. The frequency of the dynamic noise in our study was similar to the frequency used in investigating the effects of noise on dichromatic thresholds, but slower than the frequency

used in the study comparing anomalous trichromatic results. Assuming that a lower temporal frequency for the dynamic noise would also produce a relatively higher threshold in anomalous trichromats, it is possible that the dynamic noise was contribution to the larger CAD ellipses in the CVDs. A more likely factor, however, contributing to the difference in area was the fact that the CAD test measured threshold in 16 different directions that straddled the dichromat lines of confusion, whereas the CCT measured only in only 8 directions and many were located near, but not on, the tritan lines of confusion in x,y space. A larger number of samples near the regions where thresholds are the highest should produce a larger ellipse. That is, ellipse would be larger when the target colour is straddling the subject's lines of confusion. Fatigue or attention could be other factors because the CAD took twice as long to complete.

The CVN CAD ellipses in this study were, on average, larger than the average of the elliptical area calculated from Jennings and Barbur study using 4 subjects (2010); however, their elliptical areas were within our CVNs range. The median CCT major ellipse length for the CVN subjects (in u' v' space) was almost equal to median major length reported by Paramei (2012) for the same age group.

The vector length measurements of the RCCT, LandC and CCT Tri could also be used to estimate chromatic thresholds, but only in three directions near each of the dichromat lines of confusion with the grey background. The median CVN red-green thresholds were the largest on the CCT Tri followed by the RCCT, and then the LandC. The reason for the largest threshold on the CCT Tri could be the lower luminance and the luminance masking

noise on the monitor used in the test. This was different from the LandC and RCCT tests as neither test employed luminance noise, and the luminances of the stimuli were not equal to the background luminance at the higher chromatic contrasts (Rabin et al., 2011). The lack of a floor effect on LandC L- and M thresholds is likely the reason for lower red-green CVN thresholds relative to the RCCT.

The median red-green thresholds for the CVDs showed a different ordering of tests. The CCT Tri still had the highest thresholds but the LandC thresholds were next with the RCCT having the lowest red-green threshold. This was regardless of which red-green vector and red-green colour defective group. Again, the lower luminance and luminance masking noise on the CCT Tri display could be the reason for highest threshold. There are three possible reasons for the difference between the CVDs red-green thresholds on the RCCT and LandC. The first could be that the angular size on the LandC was slightly smaller than the RCCT. Although the difference in size was not large, the smaller size of LandC detail could be partially responsible for the higher thresholds. The second reason could be that the subject did not maintain alignment with RCCT test screen and their head movements provided luminance or hue clues. At the beginning of the trial, subjects were instructed to align the monitor by minimizing motion parallax in order to eliminate any changes in the stimuli that occurred as the subject's angle of view changed. However, their head position was not fixed and so it was possible that they were moving their head slightly in order to use the luminance and hue changes that occur on the monitor as the angle of view is changed. This was likely unique to the laptop monitor used to display the RCCT because the appearance of the LandC

129

stimulus did not change with small head movements. The third reason could be the different psychophysical procedures used to find the threshold. The RCCT used a pseudo-staircase procedure in which the changes in contrast were relatively large. The LandC used an adaptive thresholding procedure. The coarser step size used in the RCCT may result in lower thresholds. This effect of the different procedures would only be apparent when the median threshold for the LandC is above the minimum contrast displayed on the RCCT. This is why it would become apparent for the CVD results, but not the CVN results because the median LandC CVN threshold for the L and M cones was below the minimum contrasts presented on the RCCT.

Both the RCCT and LandC were very good at separating protans from deutans based on the L-cone and M-cone thresholds. Recent work showed that the RCCT and LandC had a perfect agreement with the Rayleigh matching results in terms of classifying the defect as protan versus deutan for the individuals who failed the tests (Hovis and Almustanyir, 2017). Only one protanope had an L cone vector length within the normal range on the RCCT and three deuteranomalous had an M cone vector length within the normals on both RCCT and LandC. The CCT Tri test, however, did not perform as well. Six deuteranomalous subjects were misclassified because their protan vector length was longer than their deutan vector length. A number of protan subjects were also misclassified on the CCT Tri test. Eighty-two percent (17 protanopes and 6 protanomalous) of the protans were classified as deutan because their deutan vector threshold was greater than their protan threshold. These results indicate that the test cannot diagnose the nature of a red-green colour vision defect correctly.

The reason for why this result occurred is uncertain. It could be due to CCT Tri software problem if the subject had thresholds near or beyond the monitor's colour gamut. Nearly all of the misclassified subjects had a very severe defect and they could have had thresholds near or beyond the monitor's gamut for both the deutan and protan vectors. If the monitor colour gamut was larger, then the test may have made the correct diagnosis.

The other aim of this chapter was to determine whether the chromatic thresholds measured with the newer tests would agree with Rayleigh diagnosis based on the range and mid-point. The CAD and CCT ellipse area and angle provide an excellent agreement with the Rayleigh diagnoses as to whether they have normal colour vision or a red-green defect. Using the elliptical angle as the parameter for CVN or CVD, the agreement of both tests with the Rayleigh diagnosis was exactly the same (AC1 =0.92). The discripancies were 5 deuteranomalous subjects on each test who were classifed as CVN based on the angle.

The CAD was also good at separating CVNs from CVDs using the elliptical area. The discrepancies on the CAD were one CVN who was classified as CVD and 3 CVDs who were classified as CVN. The CVN subject was a 51yrs old male who had a CAD elliptical area beyond the normal cut-off, but his angle that was within the normal range. Although the larger elliptical area was expected based on his age (Barbur and Carmona, 2015), his SNU values for both the red-green and blue-yellow direction were outside the age-matched CVN range (Barbur and Carmona, 2015). He had normal thresholds on LandC and CCT tests. The misclassified CVDs on CAD were three deuteranomalous subjects who had elliptical areas

131

and angles within the normal range. These individuals had normal values on all the other tests used in this study except for the Rayleigh midpoint.

The agreement between the Rayleigh classification and CCT test was lower due to 2 CVN subjects who were classified as CVDs and 7 CVDs were classified as CVN. The misclassified CVNs were a 22 years old female and a 51 yrs old male (different from the just mentioned subject) who had an elliptical area within the CVD range; however, both subjects had an elliptical angle within the CVN range. It is possible that larger ellipse for this male was a result of normal aging because the length of his major axis on the CCT was slightly higher than the median value for the same age group reported by Paramei (2012). However, it is interesting that only CCT showed the effects of normal aging. His thresholds on the CAD and LandC tests were within the range of the younger adults. The misclassified CVD subjects on the CCT were 7 deuteranomalous subjects who had an elliptical area within the normal range (3 of them also had normal areas on the CAD); however, 3 of these individuals had an elliptical angle within the CVDSs. This supports the practice of using both the angle and area to determine whether the person has normal colour vision or a red-green defect, particularly for the CCT.

The area for both tests had no significant correlations with Raleigh matching range in the CVN group. This was likely because the Raleigh match range varied from only 1 to 10 units and the CAD and CCT ellipse areas were relatively constant across the CVN subjects. However, the area for the CVDs on both tests did show significant correlations with the Raleigh matching range but the correlations were moderate for CAD and fair on the CCT.

The other parameter of interest was whether the elliptical angle was able to classify the red-green defect correctly as protan versus deutan. Based on the calculated AC1 coefficient agreement the CAD classification was better than the CCT, and the difference was significant. This was due to a large number of protan-deutan misclassifications on the CCT test rather than a misclassification as normal vs defective. The reason for the larger misclassification on CCT was that the test measured the threshold among 8 directions and did not straddle both the deuteranopic and protanopic lines of confusion compared with the CAD test that sampled 12 directions near the red-green dichromatic lines of confusion. The coarser sampling of the CCT would likely result in poorer separation of the protans and deutans based on the angle. The CCT Tri also performed worse in classifying the defect as protan vs deutan. This was because of the relatively large number of protan-deutan misclassifications. The AC1 level of agreement for the CCT Tri and Rayleigh match was significantly lower than the CCT elliptical angles. This indicated that measuring the discrimination ellipses, perhaps with more than 8 vectors, would be preferable to the CCT Tri in testing colour vision.

The AC1 coefficients agreement of the CAD classification (protan versus deutan) with the Rayleigh matching results based on either the ellipse angle or the diagnosis given by the program (in SNU) were the same. However, the discrepancies were different. One deuteranomalous subject was classified as protan using the ellipse angle and a different deuteranomalous was an unclassified red-green defect using CAD program in SNU.

133

For the tests that measure vector length, there were few significant correlations between the three tests and the Rayleigh match range among the CVNs and CVDs. The exceptions were that the deutans' matching range results had moderate correlations with the M-cone on the RCCT and LandC tests. The reason for the lack of significant correlations with CVN was the range of values for both the Rayleigh settings and thresholds was small. The reason for the lack of significant correlations with the protan matching range was that there was a large proportion of protanopes and so there was a ceiling effect on the Rayleigh match with majority of protans (68%) have a range of 70. One of the reasons for the lack of correlation between the CCT Tri deutan vector length and the matching range was that 23% of the deutans with the larger threshold had Rayleigh matching ranges between 6 and 13. Individuals with matching ranges with these values would be classified as having a mild to moderate defect (Cole et al., 2006; Squire et al. 2005).

The S-cone (tritan) vector length results showed a significant difference across tests for all groups. The median vector length of the LandC was the largest followed by the CCT Tri and the RCCT had the lowest threshold for both CVNs and CVDs. This rank order was different from the red-green threshold results, where the LandC median threshold was the lowest. The lower luminance and luminance noise used in the CCT display could explain why its threshold is higher relative the higher the RCCT, but then one would expect the CCT threshold to be higher than LandC as it was for the red-green vectors. The difference in psychophysical procedures and the slightly smaller size of the LandC could be responsible for the difference. As discussed previously for the CVD red-green results, the psychophysical

134

procedure used by RCCT could produce lower thresholds relative to the LandC. This effect was apparent for the S-cone data for both groups. The LandC thresholds were above the minimum S-cone contrast presented by the RCCT. The slightly smaller LandC stimulus size may have produced a slight additive affect to the psychophysical procedure. The difference in the psychophysical procedure, however, does not explain why the LandC S-cone threshold was above the CCT Tri threshold. The smaller size of the LandC stimulus could be one factor, but the RCCT stimulus was also smaller and so its threshold should also be higher than the CCT Tri. More likely, there are several factors contributing to the differences in the thresholds. The smaller size of the LandC has more of an effect on the threshold than the luminance difference and noise in the CCT Tri stimulus. The lower threshold of the RCCT was due to the psychophysical procedure with the additional possibility that subject moved their heads slightly in order to use the luminance or hue artifacts that were more obvious on the RCCT monitor.

Making a colour match using the Moreland equation was challenging for all the subjects because they could never make the two fields look identical due to the difference in saturation. Although the CVN results were within the normal range for their age group, the CVDs needed more green relative to the CVN group in order to make a match (Rüfer et al., 2012). We do not believe that this result was related to the L or M cone photopigment shift that occurs in the anomalous trichromats, because nearly half of the dichromats had this issue. We do not think that the result was due to these CVDs also having a blue-yellow defect along with a red-green defect because there were no significant correlations between the

135

LandC, CCT Tri and CAD tritan/S-cone thresholds and the Moreland matching range. Likely, the difference was due to the criterion used by each subject group to decide when a match occurred when they could only match two of the three perceptual attributes of colour; likely hue and brightness. For some reason, the CVD required more green in the mixture.

5.6 Conclusion

This chapter compared the chromatic discrimination thresholds measured by newer computer-based tests using a common colour vision space. Chromatic discrimination ellipses measured with the CAD were larger than the ellipses measured using the CCT. This difference was likely due to the different number and spacing of the chromatic vectors, fatigue, or attention as CAD took longer to complete than CCT or a combination of both factors. The higher CVN and CVD L- (protan) and M- (deutan) thresholds on the CCT-Tri relative to the LandC and RCCT could be due to the lower luminance and luminance masking noise used in the CCT display. In addition, because there was no floor effect on the LandC, the L- and M- thresholds, these thresholds were lower than RCCT and CCT Tri thresholds for CVNs. The lower L and M cone vector length on the RCCT among the CVDs might be due to the differences in psychophysical procedure combined with head movements that produced luminance or hue clues for identifying the target. As expected, S-cone thresholds did not vary between CVN and CVD. Nevertheless, the thresholds between tests did differ with the largest threshold on the LandC followed by the CCT Tri and the RCCT was the smallest. The larger vector length on the LandC might be because of the smallest

136

angular size of the stimulus detail, difference in psychophysical procedure or a combination of both. Monitor artifacts combined with slight head movements could be the reasons for the RCCT having the lowest S-cone threshold.

Chapter 6

Colorimetric Analysis of the Farnsworth and ColorDx D15 Color Vision Tests

6.1 Introduction

In Canada, several companies and agencies use the Farnsworth D15 (F-D15) to determine whether an applicant who has a colour vision defect has the adequate colour vision discrimination to perform a job safely and efficiently. As mentioned in the method section, the F-D15 consists of 15 coloured Munsell paper samples placed in small moveable cylindrical caps. The subject's task is to arrange those colour caps according to hue starting from fixed reference cap. The colours are approximately equally spaced around the hue circle, and the difference between adjacent colours is considered as an easily noticeable difference (Farnsworth, 1947).The ColorDx D15 is a new computerised program version of the F-D15. The program requires the subject to drag the coloured circle up to the top of the screen in order to use that colour to "fill" one of the empty rectangles. The colour selected should be the one that is most similar to the previously filled rectangle. Colours in the rectangles can be rearranged by clicking and dragging the colour to a different rectangle. Once the colours moved to the rectangles, however, the subject cannot return it down to the circles.

Errors on the D15 tests can vary from transpositions in which a cap or rectangle are placed only one to two positions from the correct arrangement, to major crossings in which caps or rectangles from opposite positions of the hue circle are placed adjacent to one

another. These major crossings parallel the lines of confusion for one of the three different types of dichromats (Farnsworth, 1947).

The purpose of this study is to evaluate the F-D15 and ColorDx D15 colour vision tests in terms of how well the actual dichromat arrangements match the predicted arrangements.

6.2 Methods

6.2.1 Colorimetric Measurements

The colorimetric properties of the F-D15 and the colour circles and rectangles of the ColorDx D15 were measured using PR-670 Spectroradiometer (Photo Research, Syracuse, NY). The F-D15 was illuminated with an Illuminant C fluorescent lamp. The caps were placed on a table with illuminance on the caps in the horizontal plane of 1400 lux (±5%). As recommended by the manufacturer, the photometer was turned on 15 minutes to warm up before taking any measurements. The photometer was properly aligned and focused on the caps using the eyepiece. The distance from the photometer to the caps was approximately 75 cm. The angle of the photometer with respect to the cap was approximately 45°. The measurement aperture diameter was 0.5^o with an exposure time of 10 msec. The measurement speed was normal, and the average of three measurements was taken for each reading. The final results were an average of 3 readings. A white diffusing plate was used as the white reference.

139

The ColorDx D15 was displayed on a Microsoft Surface Pro (Model number 1631) with Windows 10 operating system. The monitor was calibrated using Spyder 4.5.4 colorimeter to a white reference of 6500° K correlated colour temperature. During the measurement, the test was placed on a table on a dim (1 lux) room lighting. The angle of the photometer with respect to the test's screen was approximately 90° with a distance of approximately 75cm. The other measurement conditions were the same ones used for the F-D15. The Surface Pro white screen was used as the reference white by setting the RGB values to 255, 255 and 255.

6.2.2 Colour Differences

The colour differences between the colour caps (F-D15) and rectangles/disks (ColorDx D15) were calculated using the CIE L*a*b* chromaticity space for normal and dichromat colour vision. The L* corresponds to brightness, a* corresponds to a red-green dimension, and b* corresponds to a blue-yellow dimension (Wyszecki & Stiles; 1982). The nonlinear CIE L*a*b* was selected over a linear transformation in order to facilitate comparisons with previous experiments (Ramaswamy & Hovis, 2011). Appendix 1 shows the equations used to calculate the colour differences in normal colour space.

6.2.3 Colour Differences in the Dichromat Space

The dichromat colour differences were calculated based on the procedure used by Ramaswamy and Hovis (2011). Note that only the red-green dichromat spaces were used in this study. Briefly, the tristimulus values were converted to L, M, and S-cone responses using

Golz and Macleod equation 5 (Golz and MacLeod, 2003). Next, the colour-normal cone

responses were converted into deuteranopic and protanopic cone responses using the Brettel

et al. (1997) algorithm. These cone responses were then used to calculate the dichromat

tristimulus values and then the dichromat L*a*b* colour difference. Appendix 1 shows the

various steps used to calculate the colour differences in dichromat space.

6.3 Predicted Order

The ΔEs for all possible comparisons, which includes the luminance, (Tables A, B, and in

Appendix 2) were used to predict the possible order of the two D15 tests using a procedure

that was similar to the actual test instructions. That is, the order was set so that the ΔE

between adjacent caps was minimized for each type of observer. Nevertheless, it is possible

that a cap with a larger ΔE will be a viable option because the ΔE could be below a

dichromat's threshold and so there were other possible choices for ordering. For this reason,

additional predictions were carried out for these situations when the next cap had a ΔE

greater than the minimum, but less than 6. The subsequent caps were placed based on

minimum ΔE. The value of 6 was selected after calculating the average difference between

adjacent lines of confusion in dichromat space for the F-D15 test. The average difference

between the lines of confusions was a ΔE of 7, which corresponds to one just-noticeable-

difference (JND) for a dichromat. A ΔE of 6 represents an integer value that was less than the

average of dichromatic JND.

6.4 Subjects

The study used the 60 CVN subjects and 26 red-green dichromat subjects (CVD). Based on the Rayleigh colour match using the HMC Oculus anomaloscope, there were 7 deuteranopes and 19 protanopes. The mean age for the CVN was 26.3 (SD± 9.4), and 28.1 (SD± 11.5) for the dichromats. Additional information regarding the subjects was described in section 3.1.

6.5 Testing

In this study, the F-D15 and ColorDx D15 tests were used. Section 4.1.4 describes the testing procedures for each test. Both the predicted and actual D15 results were analyzed by counting the number of crossings and transpositions. A major crossing was defined as a difference between adjacent cap numbers that was greater than 2. The results were also analyzed using the Vingrys and King Smith (1988) Colour Differences Vectors analyses (CDV). The three parameters of interest were the C-index, S-index and angle. As mentioned previously, the C-index indicates the severity of the defect, and it is correlated with the number of crossings. The S-index provides a measurement of how regularly the crossings are oriented. A low S-index value indicates that there is a high degree of randomness in the arrangement. The angle gives measurement as to the type of defect (Vingrys and King-Smith, 1988). All of these parameters were calculated for each arrangement using a custom GNU Octave program (version 4.0.0, 2015).

6.6 Results

Because the colours on the ColorDx D15 were displayed as filled circles in the middle third of the screen or filled rectangles at the top of the screen, it is possible that the colours at the two locations differed. Table 10 shows the average difference between the circle and rectangle in each space along with the range. The ΔE includes the luminance term. The ΔE in any of the three colour spaces never exceeded 2.14, and the predicted order using the minimum ΔE criterion was the same for the circles and rectangles. Although the colour differences between the rectangles and circles were small, the rectangle data were used to predict the arrangements because the subjects were allowed to rearrange the rectangles as needed.

Table 10. The colour difference (ΔE) between the circles and the same rectangles for Surface-pro ColorDx D15 on the normal, protanopic and deuteranopic colour spaces.

	Normal		Protanope		Deuteranope	
	Average	Range	Average	Range	Average	Range
ColorDx D15	1.13	0.39 to 2.17	1.1	0.27 to 2.12	1.1	0.25 to 2.14

Figure 45 shows the F-D15 and ColorDx D15 chromaticity coordinates in normal a* b* space. The colour difference between each cap of the two tests is listed in the parentheses. This colour difference is based on only a* and b* coordinates because the ColorDx D15 was uniformly brighter than the F-D15. The ΔE average between both tests for each cap was 1.93 with a range of 0.35 to 3.47. For reference, a ΔE of approximately 2 represents a just-noticeable-difference in colour normal space if the coloured samples are placed adjacent to

one another (Brainard, 2003). Although there is a shift of the ColorDx D15 colours toward

green, the small ΔEs between the two tests suggest that the tests should have the same level

of difficulty.
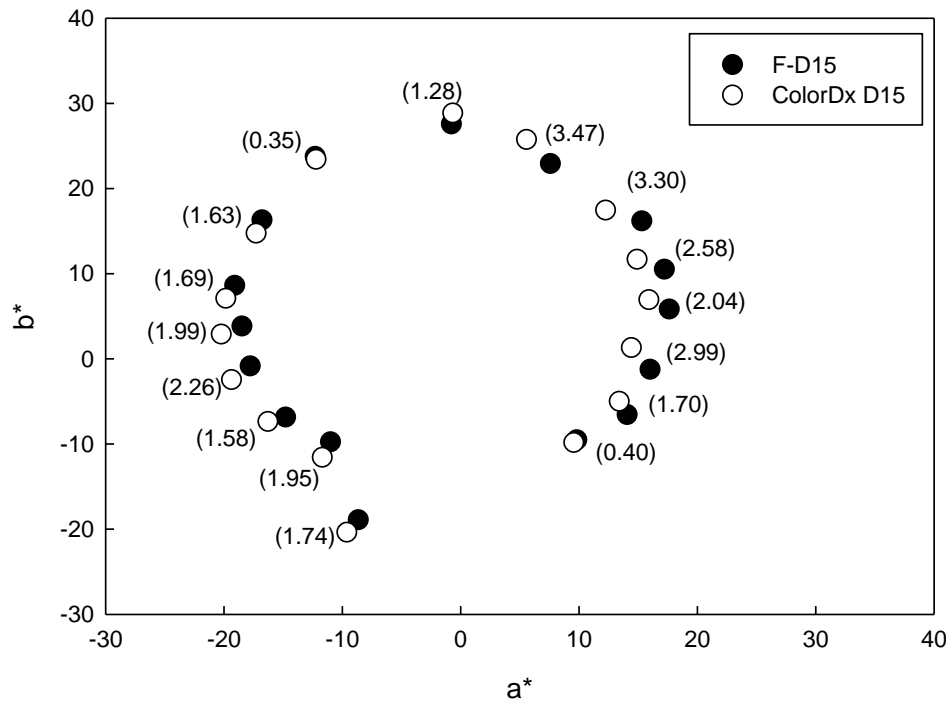
Figure 45. Chromaticity plot of a* and b* in normal space for F-D15 and ColorDx D15 tests. The numbers in parentheses are the colour difference between the ColorDx D15 and F-D15 individual caps. The colour difference is for the a* and b* dimensions. The solid circles are the F-D15 caps, and the open circles are the ColorDx D15 rectangles.

The predicted order showed that the colour-normals should have a perfect

arrangement of the caps for both tests. Ninety-two percent (n=55) of the CVN had a perfect

arrangement on the F-D15 and 98% (n=59) had a perfect arrangement on the ColorDx D15. The 5 CVNs subjects had only one transposition on the F D15, which was one of the following: switching caps 1 with 2 (ΔE=5.08), 5 with 6 (ΔE=8.15), 11 with 12 (ΔE=4.71), 13 with 14 (ΔE=5.70), and 14 with 15 (ΔE=5.20) on the F-D15. The one transposition on the ColorDx D15 was switching rectangles 11 with 12 (ΔE=4.85).

Figure 46 shows the parameter results of the predicted and the actual dichromat results for the F-D15 test. The data for the predictions is the average of 63 possible arrangements for the protanope and 384 arrangements for deuteranopes. The predictions using the only minimum ΔE between adjacent caps are shown separately for comparison. In general, the mean values based on the predictions were close to, and overlapped with, the actual dichromat results for the crossings, transpositions, and the CDV parameters. Although, the ranges of the predicted orders angles overlapped with the deuteranopic results, the mean predicted angle and angle using the minimum ΔE criteria were slightly more negative than the subjects' results. The mean predicted S-indices for both deuteranope and protanope were slightly lower than the dichromat results, but the predicted S-indices were within the range of the actual results. The predicted S-indices using the minimum ΔE, however, were almost the same as the dichromat results.
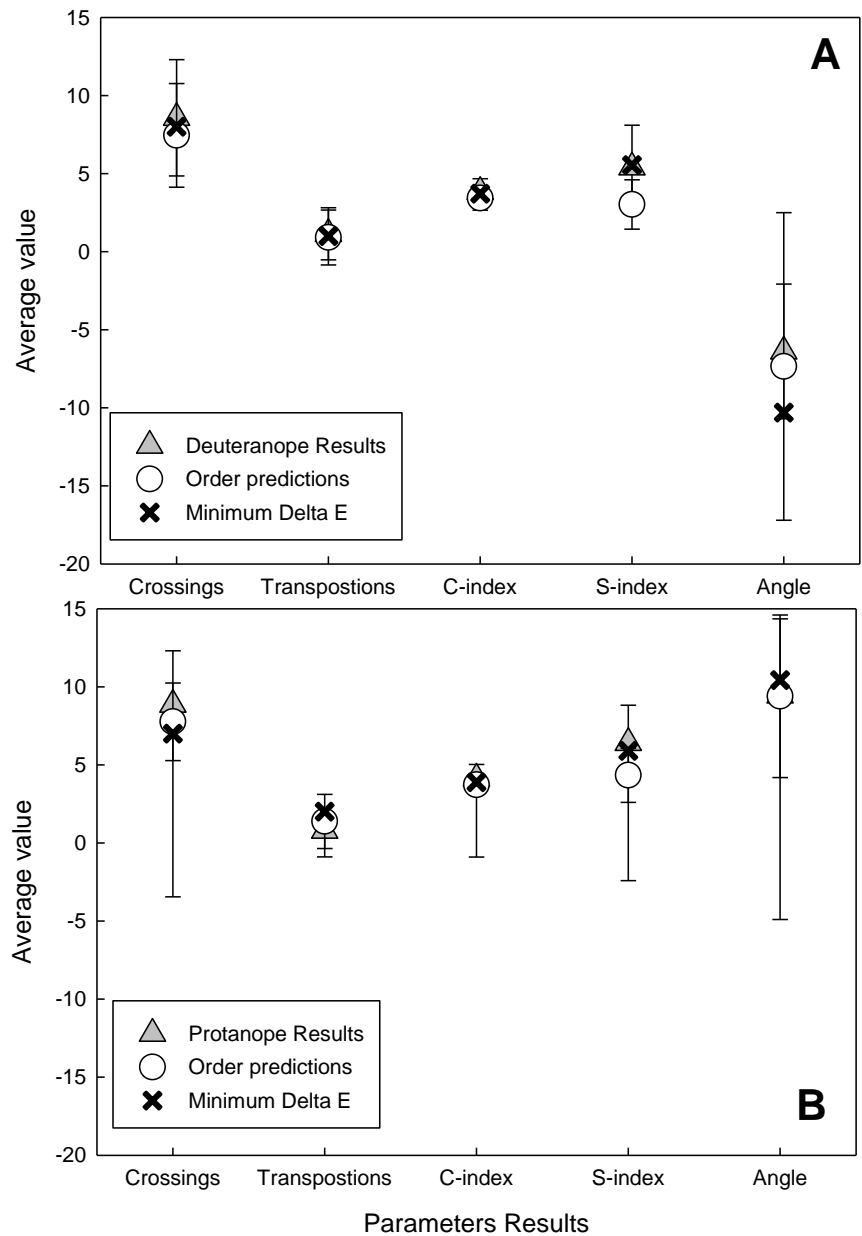
Figure 46. Average of the predicted (open circles) and dichromatic arrangement (grey triangle) and the minimum ΔE between adjacent caps arrangement (black cross) results for the F-

146

D15. A is the deuteranopic results and B is the protanopic results. The errors bars are the +2 standard deviations of the mean for both the predicted and dichromat results.

Figure 47 shows the ColorDx D15 predictions and dichromats results. The predicted results represent the average of 207 possible arrangements for the protanope and 280 different arrangements for the deuteranope. In general, the predicted results were very similar to the dichromat results. The exceptions were the higher number of predicted crossings, S-index for the deuteranope using the minimum ΔE criterion and the more positive predicted mean angle and lower mean S-index for the protanopes.
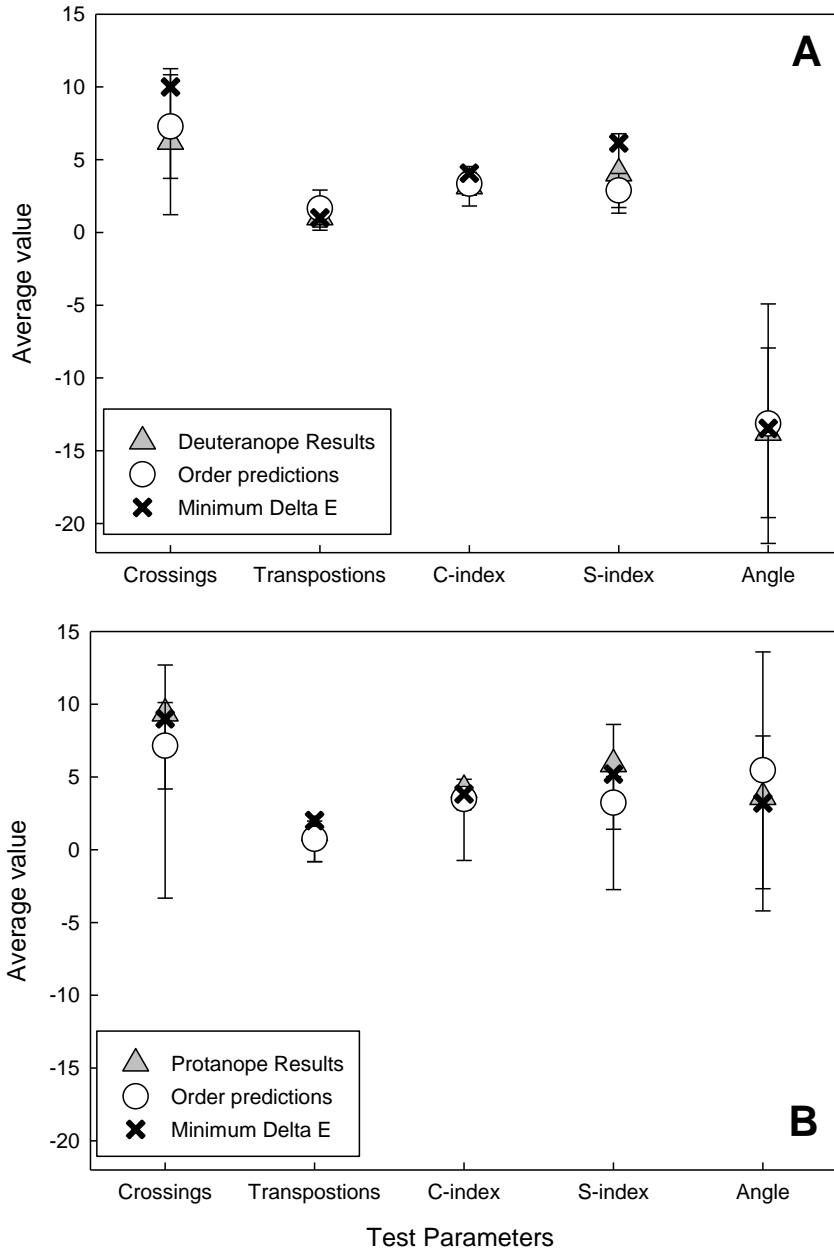
Figure 47. Average of the predicted (open circles) and dichromatic arrangement (grey triangle) and the minimum ΔE between adjacent caps arrangement (black cross) results for the

ColorDx D15. A is the deuteranopic results and B is the protanopic results. The errors bars are the ±2 standard deviations of the mean for both the predicted and dichromat results.

The order predictions across tests were that the number of crossings and C-index for each dichromat should be equal on each test, the dichromat angles on the ColorDx D15 should be more negative (or less positive) than their F-D15 values. The predictions for S-index were that dichromat values should be lower on the ColorDx D15. The deuteranope subjects actually had fewer crossings on the ColorDx D15 and marginally more crossings than predicted on the F-D15, whereas the protanopes had more crossings than predicted on both the F-D15 and ColorDx D15. Because the number of crossings and C-index are correlated (Atchison et al.1991), the predicted C-indices of each test paralleled the differences in the predicted number of crossings. The dichromat subjects' angles on the ColorDx D15 were more negative as predicted. Nevertheless, the predicted angles for both tests were within limits suggested by Vingrys and King-Smith to diagnose the type defect.

6.7 Discussion

This study showed that the dichromat transformations of the D15s colours in L*a*b* space provided predictions that were very similar to the actual dichromat arrangements. Although the predicted arrangements using the minimum delta E between adjacent caps were within the range of the actual dichromat arrangements, they were not always as good as the mean values when other possible arrangements were considered. Allowing alternative arrangements when two caps may be within 1 JND, obviously increased the number of possible arrangements, but it also provided a better fit to the average data and estimate of the

possible variability in actual arrangements. Our model suggests that 1 JND in dichromat

space corresponds to a dichromat ΔE of greater than 6 in the respective red-green dichromat

colour space, at least for tasks similar to the D15's. Although, a ΔE of 2.2 is considered to be

a 1 JND in colour-normal L*a*b* space, 10% of the colour-normals may require a ΔE of

4.71 based on our D15 results.

There were, however, some discrepancies between the predicted order and the actual

arrangements. The first was that the predicted S-index was lower than dichromat results for

all conditions, but especially for the F-D15. This indicates that predicted arrangements

tended to deviate from a regular dichromat pattern and towards a more random arrangement.

This occurred because the alternative arrangements resulted in situations where the

remaining few caps had relatively large ΔEs between them and the last cap placed in order.

Placing these "left-over" caps at the end resulted in irregular arrangements or mixed tritan

and red-green arrangements. This would be similar to a patient having a few caps left over

that do not really fit with the last one put in the box, but they cannot decide what to do with

them and just place them at the end.

The second discrepancy was that the predicted number of crossings for the protanope

on the ColorDx D15 was lower than the actual results and lower than using the minimum ΔE

criterion. This suggests that our criterion of a ΔE of 6 may be too lax from the protanopes, at

least on the ColorDx D15. As to why the protanopes behaved differently on the ColorDx D15

is not certain, but it may be related to the slightly larger angular size of the ColorDx D15

rectangles and circles, luminance artifacts or both. The third discrepancy between the

150

predicted and actual arrangements was the predicted protanopic angle on the ColorDx D15 was more positive than the actual results. The more positive angles were likely a result of allowing multiple possible arrangements. For several cases, the arrangement of the last few caps/rectangles resulted in somewhat random/tritan appearance when plotted on the score sheet. These arrangements had more positive angle and larger S-index, which increased the average of both of these values for the predicted arrangements. This suggests that the ΔE criterion for a JND may have been slightly large.

Both the predictions and data showed that there are differences between the F-D15 and ColorDx D15. First, both the predictions and actual arrangement showed that deuteranopes had fewer crossings on the ColorDx D15. Second, there was a negative shift in the dichromat angles on the ColorDx D15 that was both predicted and obtained in the results. Figure 45 shows that the negative shift is likely due to the counter-clockwise rotation of most of the ColorDx D15 colours relative F-D15. The negative shift suggests that changing the minimum value of the protan angle from zero to -3 as suggested by Atchison et al. (1991) for the F-D15 may also apply to the ColorDx D15.

6.8 Conclusion

A colormetric evaluation of the F-D15 and ColorDx D15 tests showed good results in predicting dichromat and CVN arrangements on both tests. Both the predictions and results showed that deuteranope subjects had fewer crossings on the ColorDx D15 test, which

151

suggests that the ColorDx D15 would be easier to pass for deutans. However, protanope

subjects had more crossings than the predicted on both F-D15 and ColorDx D15.

Chapter 7

Comparison of the ColorDx D15 with the Farnsworth D15 Colour Vision Tests using two Scoring Criteria

7.1 Introduction

The Farnsworth-Munsell D15 (F-D15) colour vision test was introduced to divide normal and those with a mild colour vision deficiency from those individuals with a moderate-to-severe colour vision deficiency (Birch, 2008). In Canada, it is used to exclude individuals from occupations such as policing, firefighting, and electronics who do not have adequate colour discrimination to carry out their work safely and efficiently. The test is designed based on colour confusions for protan, deutan, and tritan. The ColorDx D15 test is a new computerised colour vision test that is similar to the F-D15 test in design. Both tests are usually analysed by counting the number of crossings and transpositions. A major crossing is usually defined as a difference between adjacent cap numbers that is greater than 2. The transposition occurs when a cap, or rectangle, is placed only one to two positions from the correct arrangement.

In Canada, many agencies have the further stipulation that the applicant has to pass on 2 of the 3 trials in order ensure that a single passing arrangement is repeatable. Birch (2008) has proposed another scoring criterion that would reduce the probabilities that dichromats pass the test. Her recommendations are

- Any major crossing is a failure.

- Retest the applicants who had only two major crossings to determine whether the results are repeatable.

- Allow the subject to review the first arrangement to increase the pass rate when there are a small number of crossings in the arrangement.

Although the 2 out of 3 trials offers a method for assessing the within-session variability, the repeatability of the F-D15 on different days using this criterion has never been assessed to our knowledge. The purpose of this chapter is to determine:

o The repeatability within a session (1st trial vs 2 out 3 trials) of the ColorDx D15 and F- D15 in performing the tests without feedback using Farnsworth (1947) and Birch (2008) failure criteria.

o The repeatability of the tests for the two different crossing criteria on different days using the first trial only and 2 out 3 trials.

o The repeatability of classification on different days for the 1st trial only and 2 out 3 trials using Farnsworth and Birch criteria for both ColorDx D15 and F-D15.

o The validity of ColorDx D15 relative to the F-D15.

7.2 Methods
The F-D15 and ColorDx D15 test procedures are described in section 4.1.4.

7.3 Subjects
Sixty CVNs and 68 CVDs performed both D15 tests. All the information regarding the subjects was described in section 3.1.

7.4 Results

7.4.1 Analysis

First, we investigated whether there was a difference in the pass/fail rates using the results from the first trial (1st) versus the 2-out-of-3 trials (2/3) at the first session for each crossing criteria. A major crossing was defined as a difference between adjacent cap/rectangle numbers that was greater than 2. Note that in this study we used the kappa (κ) coefficient of agreement recommend by the Working Group 41 (1981) so that we could compare our findings with the findings in the literature. Only the CVD subjects are included in these comparisons because only individuals who fail the colour vision-screening test are required to take the F-D15. Nevertheless, all CVNs passed both tests at both visits using both crossing criteria. In terms of the classification as protan versus deutan, visual inspection of the 1st trial and 2/3 trials was carried out by a third person who was familiar with the test but was masked as to the anomaloscope results. The classification categories were protan, deutan or unclassified. The mixed category was used when it was not clear whether there was a protan or deutan defect (i.e. equal number of crossings in each direction). The evaluation order of the classification were the F-D15 1st trial only and then 2/3 and then the ColorDx D15 1st trial only and then 2/3.

7.4.2 ColorDx D15

7.4.2.1 Repeatability within session

Figure 48 shows the ColorDx D15 pass/fail for just the first trial results compared with pass/fail results for 2/3 trials for using Farnsworth's and Birch's failure criteria. There was a good agreement between the 1st trial and 2/3 trials with the κ coefficient agreement of 0.97 (95% CI=0.91 to 1) for both criteria. The 95% confidence interval included 1, which indicates that there was essentially perfect agreement between 1st trial-only and 2/3 trials. There was only one discrepancy for both crossing criteria. This was a deuteranomalous who passed the test at the 1st only but failed it at 2/3 trials. All but one dichromat failed the ColorDx D15 test using both crossing criteria. The exception was a deuteranope who passed the test using both crossing criteria.
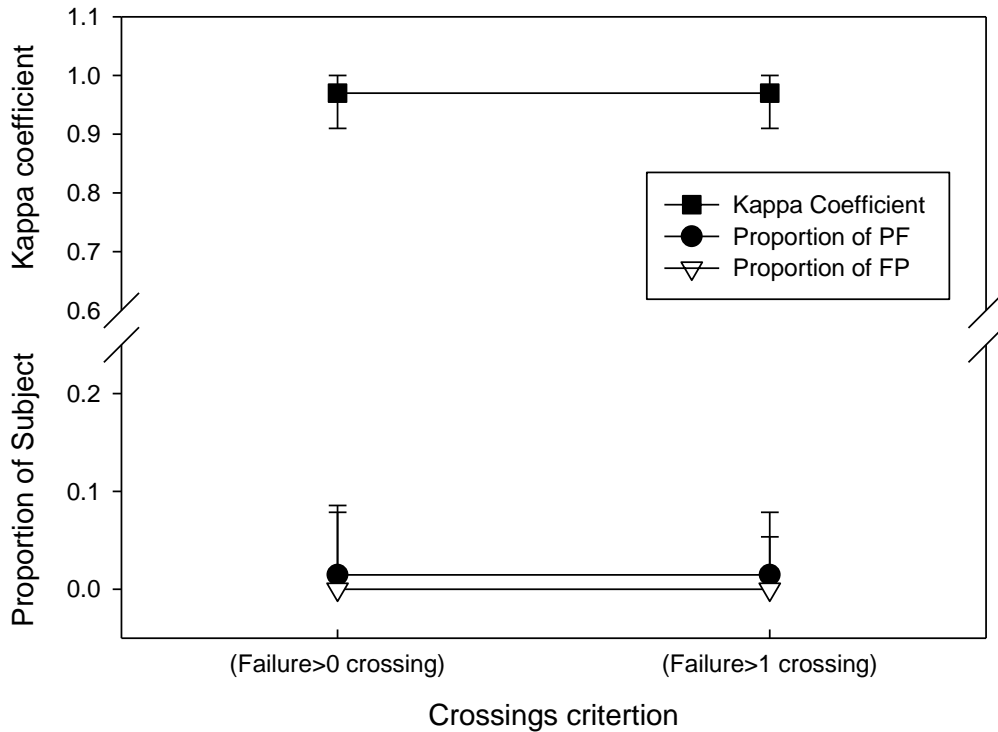
Figure 48. The ColorDx D15 within session coefficient of agreement values (top) and the frequency of within-session discrepancies (bottom) for the 1st vs 2 out 3 using Birch's (>0 crossing) or Farnsworth's (>1 crossing) or failure criteria. PF is the frequency of subjects who passed the first trial but failed at the 2 out 3 trials. FP is the frequency of subjects who failed the first trial only but passed the 2 out 3 trials. Errors bar are the 95% confidence intervals.

7.4.2.2 Repeatability between visits

Figure 49 shows the ColorDx D15 between-visit repeatability results for the 1st trial and 2/3 trial in the ColorDx D15 test using the two scoring criteria on different days. The repeatability coefficients for the 1st trial only and 2/3 trials using Farnsworth's criterion were

higher than the Birch's criterion, but only the 1st trial-only results were significantly higher based on the 95% confidence intervals. The number of discrepancies was less than 10% of the subjects for each trial-number-crossing criterion. If Farnsworth's criterion was used, then a slightly higher proportion subjects failed the first session, but passed at the second visit. All the discrepancies were anomalous trichromats. More specifically, the only between-visit discrepancy using more than one crossing as the failure criterion on the first-trial only was a protanomalous who failed at the first session but passed at the second session. The between-visit discrepancies using 2/3 trials were the previously mentioned protanomalous subject and 2 deuteranomalous individuals. All three failed at the first visit, but passed at the second visit.

For Birch's criterion, the types of discrepancies were equal for the 2/3 trials, whereas the proportion of subjects who passed the first session, but failed the second was marginally higher for the 1st trial only but the difference was not significantly different. The between-visits discrepancies when using the 1st trial for Birch's any- crossing criterion were one deuteranomalous who failed the first but passed the second session and 4 anomalous trichromats (2 deuteranomalous and 2 protanomalous) who passed the first but failed the second session.

The discrepancies for the 2/3 trials were,

- one deuteranomalous and one protanomalous who failed the first session but passed the second session. This deuteranomalous was the same subject who failed the first session but passed the second session when one crossing was allowed to pass the test.

- two protanomalous passed the first session but failed the second session. Both subjects were the same subjects who pass the first session but failed the second session using the first trial only.
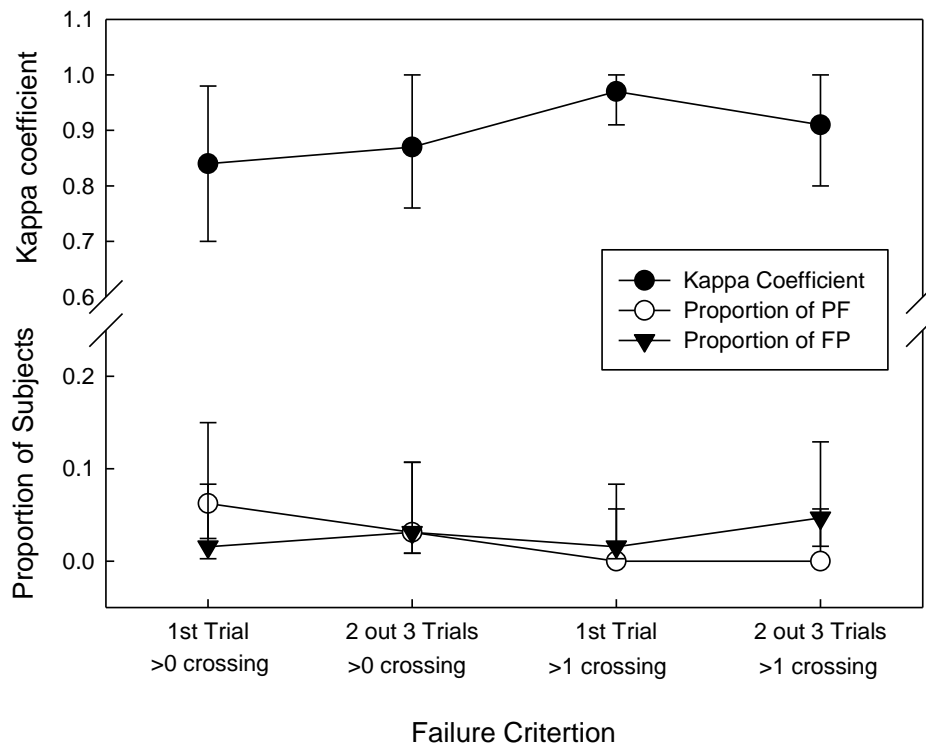
Figure 49. Between visits coefficient of agreement values (top) and the frequency of between visits discrepancies (bottom) for the ColorDx D15 using Birch's (>0 crossing) or Farnsworth's (>1 crossing) failure criterion. PF is the frequency of subjects who passed the first visit but failed the second visit. FP is the frequency of subjects who failed the first visit but passed the second visit. Errors bar are the 95% confidence intervals.

159

7.4.2.3 Repeatability of Classification

7.4.2.3.1 Repeatability of classification within session

Table11 shows the repeatability of classification within the session for the ColorDx D15 using both scoring criteria. The agreement values between the 1st trial and 2/3 trials were slightly higher, but not significantly different, based on the 95 % confidence interval when using the > 1 crossing-criterion (κ= 0.89, 95% CI=0.75 to 1) compared with no crossing criterion (κ=0.85, 95% CI=0.69 to 1). The discrepancies when using Farnsworth's criterion were one deuteranomalous who had a mixed classification at the 1st trial but classified as deutan using 2/3 trials. The other discrepancy was another deuteranomalous who was classified as protan at the 1st trial but classified as deutan using 2/3 trials. When using the Birch's failure criterion, the discrepancies were the same as the discrepancies when using Farnsworth's criterion with an additional protanomalous who was classified as deutan at the 1st trial but had a mix classification using 2/3 trials.

Table 11. Repeatability of the ColorDx D15 classification within a session.

| | | 1st trial | | | | | |
| | | Farnsworth's criterion | | | Birch's criterion | | |
		Protan	Deutan	Mixed	Protan	Deutan	Mixed
2 out 3 trials	**Protan**	21	0	0	21	0	0
	Deutan	1	14	1	1	15	1
	Mixed	0	0	0	0	1	0

7.4.2.3.2 Repeatability of classification between visits

Table 12 shows the repeatability between visits for ColorDx D15 classification using the > 1 crossing-criterion. The agreement between visits was slightly higher but not statistically significantly different, based on the 95% confidence interval when using the 1st trial only (κ=0.94, 95% CI=0.82 to 1) compared to the 2/3 trials (κ=0.87, 95% CI=0.70 to 1). The discrepancy when using just the 1st trial was one deuteranomalous who had a mixed classification at the first session but classified as protan at the second session. For the 2/3 trials, one deuteranomalous was classified as deutan at the first session but was classified as protan at the second session. This subject was the same subject who had mixed classification using the 1st trial only. The other discrepancy was a protanope who was classified correctly at the first session but was classified as deutan at the second session.

Table 13 shows the repeatability between visits of ColorDx D15 classification when any crossing is a failure. The coefficients of agreement for classification were 0.94 (95% CI=0.83 to 1) for the 1st trial only and 0.88 (95% CI=0.71 to 1) for the 2/3 rule, which were nearly equal to the previous values when more than one crossing is a failure. The discrepancies were also the same individuals.

Table 12. Repeatability of the ColorDx D15 classification using Farnsworth's criterion.

		1st session					
		1st Trial			2 out 3		
		Protan	Deutan	Mixed	Protan	Deutan	Mixed
2nd session	Protan	20	0	1	19	1	0
	Deutan	0	13	0	1	12	0
	Mixed	0	0	0	0	0	0

Table 13. Repeatability of the ColorDx D15 classification using Birch's criterion.

		1st session					
		1st Trial			2 out 3		
		Protan	Deutan	Mixed	Protan	Deutan	Mixed
2nd session	Protan	21	0	1	20	1	0
	Deutan	0	13	0	1	13	0
	Mixed	0	0	0	0	0	0

7.4.3 F-D15

7.4.3.1 Repeatability within session

Figure 50 shows the F-D15 pass/fail for just the 1st trial results compared with pass/fail

results for 2/3 trials for the CVD group using any major crossing or more than one crossing

as the failure criterion. The coefficient of agreement between the 1st trial and 2/3 trials were

0.97 (95% CI=0.91 to 1) for both failure criteria. The 95% confidence intervals for both

failure criteria includes 1 (perfect agreement), which indicate that the two methods of test

administration gave essentially identical results in terms of a pass/fail. The only discrepancy

was a deuteranomalous (different subject at each scoring criterion) who failed the 1st trial but

passed the 2/3 trials.
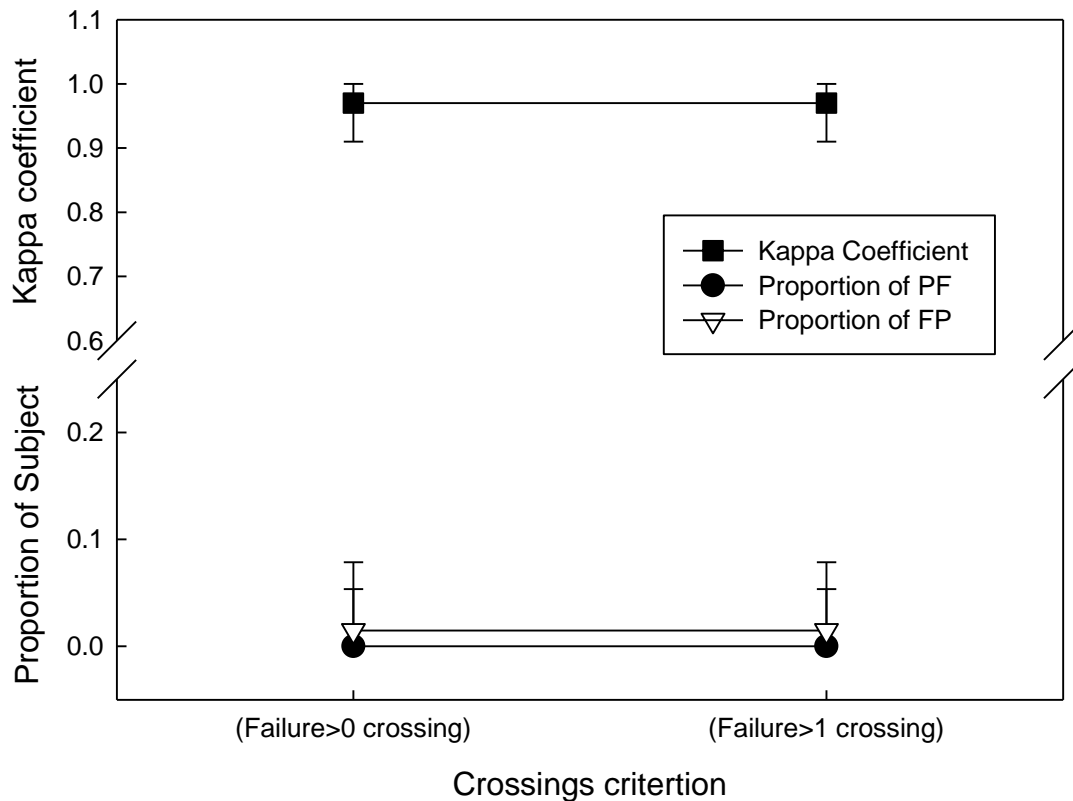
Figure 50. The F-D15 within session coefficient of agreement values (top) and the frequency of within-session discrepancies (bottom) for the 1st vs 2 out 3 using Farnsworth's (>1 crossing) or Birch's (>0 crossing) failure criterion. PF is the frequency of subjects who passed the 1st trial,

163

but failed at the 2 out 3 trials. FP is the frequency of subjects who failed the 1st trial only, but passed the 2 out 3 trials. Errors bar are the 95% confidence intervals.

7.4.3.2 Repeatability between visits

Figure 51 shows the repeatability of using the 1st trial only and 2/3 trials in the F-D15 test using the two criteria on different days. The agreement between the first session and the second session using the 2/3 trials was the best when using Farnsworth's criterion with a κ coefficient of agreement of 0.97 (CI=0.91 to 1). This value was significantly higher than both test administration options using Birch's criterion based on the 95% confidence interval. Although the coefficient of agreement for the 1st trial using Farnsworth's criterion was slightly lower than the 2/3 trials, it was not statistically different based on the 95% confidence interval. The proportion of subjects who failed the first session, but passed the second (less than 10% of the subjects) was uniformly higher across all test conditions even though the difference was small. The discrepancies using Farnsworth's failure criterion were one deuteranomalous and one protanomalous using the 1st trial only and the same protanomalous using the 2/3 trials who failed the first session but passed the second session. The discrepancies using Birch's criterion for the 1st trial only were

- Four subjects (one protanomalous and three deuteranomalous) who failed the first visit but passed the second session.
- Two protanomalous passed the first visit but failed the second visit.

164

When using the 2/3 rule, three subjects (two deuteranomalous, one protanomalous) failed on the first day but passed on the second day. These subjects were the same subjects who failed at the first session, but passed at the second session using the 1st trial only protocol. The other discrepancy was a protanomalous who passed the first session but failed the second session. This protanomalous also passed the first session but failed the second session using the 1st trial only.
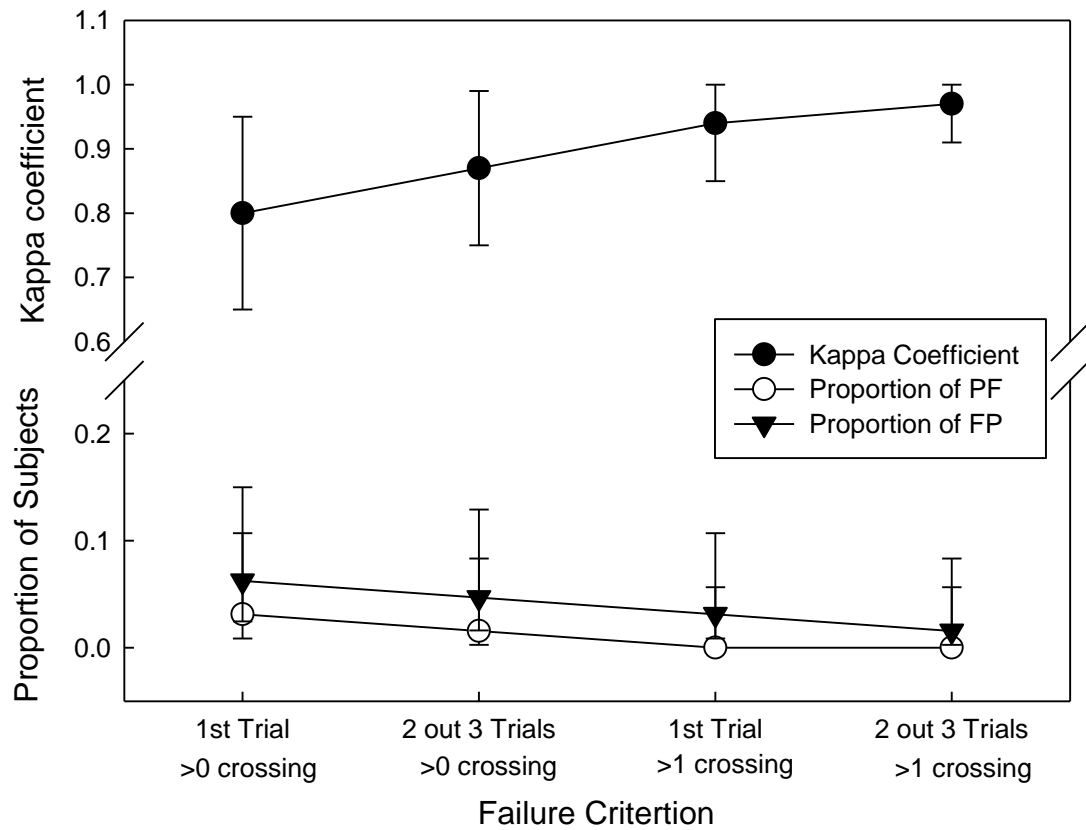
Figure 51. Between visits coefficient of agreement values (top) and the frequency of between visits discrepancies (bottom) for the F-D15 test using Farnsworth's (>1 crossing) or Birch's (>0

crossing) failure criterion PF is the frequency of subjects who passed the first visit but failed the second visit. FP is the frequency of subjects who failed the first visit but passed the second visit. Errors bar are the 95% confidence intervals.

7.4.3.3 Repeatability of Classification

7.4.3.3.1 Repeatability of classification within session

Table 14 shows the repeatability of classification within the session for the F-D15 using Farnsworth's and Birch's criteria. The classification agreements between the 1st trial and 2/3 trials were the same when using the >1 crossing-criterion or no crossing criterion (κ=0.94, 95% CI=0.75 to 1). The discrepancies were also the same with 2 deuteranomalous was classified as protan at the 1st trial but classified as deutan using 2/3 trials.

Table 14. Repeatability of the F-D15 classification within a session.

		1st trial					
		Farnsworth's criterion			Birch's criterion		
		Protan	Deutan	Mixed	Protan	Deutan	Mixed
2 out 3 trials	Protan	23	0	0	23	0	0
	Deutan	2	15	0	2	16	0
	Mixed	0	0	0	0	0	0

7.4.3.3.2 Repeatability of classification between visits

Table 15 shows the repeatability between visits of the F-D15 classification based on the visual inspection using both criteria. The repeatability of classification based on the Farnsworth's and Birch's failure criteria were the same, and so only one table was presented. For the 1st trial only result, the κ coefficient of agreement for classification on different days was 0.71 (95% CI=0.47 to 0.95) and 0.94 (95% CI=0.83 to 1) using 2/3 trials. These values were significantly different based on the 95% confidence interval. Two deuteranomalous subjects using the 1st trial only rule were classified incorrectly as protan at the first visit but were classified correctly at the second visit. One deuteranomalous and one deuteranope were classified correctly at the first visit but as protans at the second visit. The remaining protanomalous was classified correctly at the first visit, but had a mixed classification at the second visit. The only discrepancy for the 2/3 trials was the one deuteranope mentioned previously was classified incorrectly as a protan at the second visit.

Table 15. Repeatability of the F-D15 classification using either Farnsworth's and Birch's criteria.

		1st session					
		1st Trial			2 out 3		
		Protan	Deutan	Mixed	Protan	Deutan	Mixed
2nd session	Protan	20	2	0	21	1	0
	Deutan	2	11	0	0	14	0
	Mixed	1	0	0	0	0	0

7.4.4 Agreement between both tests

7.4.4.1 Pass/fail

Table 16 shows the pass/fail agreement between the ColorDx D15 and F-D15 for the 1st trial only on the CVD group at the first visit using Farnsworth's and Birch criteria. The agreement between the two tests for the 1st trial only was marginally higher when using >1 crossing as a failure criterion (κ=0.79 (95% CI=0.63 to 0.94) compared with no crossings allowed (κ=0.72 (95% CI=0.55 to 0.89). The difference was not significant based on the 95% confidence interval. Using the Farnsworth criterion of >1 crossing, 57% of the discrepancies (n=4) were protanomalous. The majority of discrepancies (55.5% (n=5)) were deuteranomalous using Birch's criterion of no crossings allowed. Although the number of protanomalous subjects was 4 who had discrepancies between the tests, 75% of them (n=3) failed the F-D15 and passed the ColorDx D15.

Table 17 shows the pass/fail agreement between the ColorDx D15 and F-D15 for the 2/3 trials at the first visit using Farnsworth's and Birch criteria. The coefficient of agreement between the two tests using both Farnsworth's and Birch's criteria was 0.88 (95% CI=0.75 to 1). This value was significantly higher than the agreement for the 1st trial only using Birch's criterion, but not Farnsworth's criterion. All the discrepancies were anomalous trichromats.

Table 16. The agreement between the ColorDx D15 and F-D15 on the 1st visit pass/fail results for the 1st trial only using both failure criteria.

1st Trial		F-D15			
		Farnsworth's criterion		Birch's criterion	
		Pass	Fail	Pass	Fail
ColorDx D15	Pass	24	4 (3DA& 1 PA)	22	6 (5 DA & 1 PA)
	Fail	3 (3 PA)	37	3 (3 PA)	37

*PA is protanomalous and DA is deuteranomalous.

Table 17. The agreement between the ColorDx D15 and F-D15 on the 1st visit pass/fail results for the 2/3 trials using both criteria.

2 out 3		F-D15			
		Farnsworth's criterion		Birch's criterion	
		Pass	Fail	Pass	Fail
ColorDx D15	Pass	27	3 (2 DA & 1 PA)	23	1 (1 DA)
	Fail	1 (1 PA)	37	3 (3 PA)	41

*PA is protanomalous and DA is deuteranomalous.

7.4.4.2 Classification

Table 18 shows the classification agreement between the ColorDx D15 and F-D15 using the 1st trial only. The agreement between the two tests was slightly better when using Farnsworth's criterion (κ= 0.83, 95% CI=0.64 to 1) compared with Birch's criterion (κ= 0.78, 95% CI=0.58 to 0.99), but the difference was not significant based on the 95%

169

confidence interval. When using Farnsworth's criterion of >1 crossing, each test classified a different deuteranomalous subject as a protan. The one mixed-protan outcome was a deuteranomalous who was misclassified as protan by the ColorDx D15. When using Birch's criterion of no crossings allowed, there were the same three discrepancies and an additional deuteranomalous subject was classified incorrectly as protan by the F-D15.

Table 19 shows the classification agreement between the two tests using 2/3 trials. For the Farnsworth's criterion, the classification agreement between tests classification was very good with a κ coefficient of 0.95 (95% CI=0.83 to 1). The one discrepancies was a deuteranomalous subject who was classified incorrectly as protan by the F-D15. This subject was the same subject who was unclassified by the ColorDx D15 and was classified incorrectly as protan by F-D15 using the 1st trial only. The agreement between the two tests using Birch's criterion was slightly lower with κ coefficient of 0.89 (95% CI=0.75 to 1), but the difference was significant from the value using Farnworth's criterion based on the 95% confidence interval. The 2 discrepancies were 2 deuteranomalous subjects who were classified incorrectly as protan by F-D15. One of them was the same subject who was classified incorrectly by F-D15 using Farnsworth's criterion for either the 1st trial only or 2/3 trials, and the other one was the same subject who was classified incorrectly by F-D15 using Birch's criterion for the 1st trial only.

Table 18. The ColorDx D15 classification agreement with the F-D15 using the 1st trial only.

1st Trial		F-D15					
		Farnsworth's criterion			Birch's criterion		
		Protan	Deutan	Mixed	Protan	Deutan	Mixed
ColorDx D15	Protan	20	1	0	20	1	0
	Deutan	1	13	0	2	13	0
	Mixed	1	0	0	1	0	0

Table 19. The ColorDx D15 classification agreement with the F-D15 using 2/3 trials.

2 out 3		F-D15					
		Farnsworth's criterion			Birch's criterion		
		Protan	Deutan	Mixed	Protan	Deutan	Mixed
ColorDx D15	Protan	20	0	0	20	0	0
	Deutan	1	16	0	2	16	0
	Mixed	0	0	0	0	0	0

7.5 Discussion

7.5.1 Within-session repeatability

Both the ColorDx D15 and F-D15 showed that there were a small number of discrepancies between the 1st trial only and 2/3 trials for both failure-scoring criteria within a session. The difference between using the 1st trial only or 2/3 trials was not statistically significant for either failure criterion, which indicates that there was no practice effect within a session for either test. Because of the small number of discrepancies, it was not possible to determine

171

whether there was an interaction between tests, criteria and type of defect (protan vs deutan). Based on just these results, it is difficult to justify using the 2/3 trial rule. However, none of the subjects was taking the test in order to qualify for a job and so the level of an anxiety was probably lower. It remains to be determined whether the additional stress of qualifying for a job affects the within-session repeatability.

7.5.2 The between visits repeatability

The between-session repeatability using either the 1st trial only or 2/3 trials was higher using Farnsworth failure criterion of more than one crossing for both the 1st trial only and 2/3 trials on both tests; however, the difference only reached statistical significance on the ColorDx D15. This suggests that using Farnsworth's criterion would be preferable because the between-session agreement was slightly higher than using any crossing as a failure, regardless of how many trials were performed within a session.

Although the number of between-session discrepancies on the F-D15 was relatively small, the general trend was that the proportion who failed the first session and passed the second was higher than those who had the opposite results for both failure criteria. This trend was also evident on the ColorDx D15 using Farnsworth's failure criterion, but it was not present when using Birch's criterion. Although the proportions were not significantly different, the results suggest that there is a subtle practice effect that may be more evident across sessions, than within session for both tests. We are uncertain as to why the protan made up the majority of the discrepancies on the ColorDx D15 even though they were the

minority of the CVD subjects. It is possible that this result could be related to how the protans may have used any luminance information that may have been present. In most cases, they used this information to their advantage at the second session.

The between-session repeatability of the ColorDx D15 and F-D15 in this study using the test recommendation of >1 crossing as a failure were approximately the same as the value calculated from Farnsworth's data who included a large number of CVNs in his study which would improve the overall repeatability (Farnsworth, 1947). However, the repeatability of both tests in this study was larger than the value (κ =0.84) calculated by Hovis et al. (2004) using the criterion greater than one crossing for only the 1st trial at each session. This difference in results is most likely due to the larger proportion of dichromats in our study compared to Hovis et al. (2004) as dichromats usually failed both sessions. In our study, none of the dichromats had a pass/fail discrepancy between visits on either test.

7.5.3 Repeatability of Classification

The classification repeatability within a session (1st trial vs 2/3 rule) on the ColorDx D15 and F-D15 using either Farnsworth's or Birch's criteria were almost the same, but the discrepancies were slightly different. There was one additional protanomalous subject who had within session discrepancies on the ColorDx D15 using Birch's criterion. This subject failed the test using Birch's criterion at the first trial, but passed the remaining 2 trials. This person passed the ColorDx D15 on 3/3 trials using Farnsworth's failure criterion.

The between visits repeatability for classification using the 1st trial-only was 94% for the ColorDx D15 and 71% for the F-D15 using either failure criterion. These values were significantly different. The 0.71 value found for the F-D15 was also lower than the 80% between-session agreement value reported by Hovis, et al. (2004) for the F-D15. Nevertheless, assuming our mixed classification corresponds to Hovis et al. (2004) and Atchison et al. (1991) mixed classification results, our rate on both sessions using the 1st trial only and either scoring criterion was identical to the 2% mixed diagnosis rate they reported (Hovis et al., 2004; Atchison et al., 1991).

There are two possible reasons for the lower repeatability for the F-D15 classification in our study. The first is that it was difficult to classify subjects who had less than 3 crossings. This was not the case on the ColorDx D15 as those subjects passed the test. It was also harder to classify individuals who had more than 8 crossings because the pattern contained a nearly equal number of protan vs deutan crossings. This also was not the case on the ColorDx D15. It does not appear to be related to the type of defect because the discrepancies were about equally distributed across the type of defect (protan vs deutan).

Increasing the number of trials per sessions, however, significantly improved the classification repeatability for the F-D15 with only a marginal change in the ColorDx D15 classification repeatability. The reason for the better repeatability in the classification with an increase in trials was likely due to the evaluator. The evaluator was familiar with visual inspection of the D15 patterns, but he had examined the 1st trials-only before the 2/3 results, and so he had substantially more experience when looking at the multiple trial arrangements.

174

There are two possible reasons for the difference in the classification repeatability between the ColorDx D15 and F-D15s. The first is related to the colorimetric properties of the ColorDx D15. The previous chapter showed that there should be less overlap in the ColorDx D15 angle for protans and deutans compared with the F-D15 and so the ColorDx D15 should be better in classification. The second possible reason could be related to the evaluator. The F-D15 (both the first and second visit) was evaluated before the ColorDx D15 and so the evaluator had more experience on evaluating the crossing on the ColorDx D15.

7.5.4 Agreement between two tests

The ColorDx D15 showed a good agreement with the F-D15 test using both failure criteria. The agreement was better using 2/3 trials compared to the 1st trial only but the difference was not statistically significant. In general, more CVDs tended to fail the F-D15 (62.5%) compared with the ColorDx D15 (59.5%). This indicates that the ColorDx D15 test would be slightly easier to pass. The previous chapter indicated that it was likely not due the different colorimetric properties since the number of crossings for dichromats is predicted to be equal for each test. It is more likely that the difference is due to the larger angular size of the ColorDx D15. Color defectives make fewer errors when the objects are larger in size (Cole et al. 2006). The difference could also be due to the fact that the ColorDx D15 colours are a metameric match to the Munsell papers and so the colours may not be a metameric match to the anomalous trichromats. Although this could be a possible explanation, the result that the

discrepancies between the tests were not large suggests that the colours were also close to a metameric match for most anomalous trichromats.

Our failure rate for the F-D15 was higher than the 54% rate reported by Birch (2008). The difference was likely due to different subject pools. Her sample had a lower percentage of dichromats who usually fail the F-D15. One of the reasons for Birch's recommendation of any crossing as a failure was that reduces the probability that a dichromat will pass the F-D15. Although it may reduce the probability of a dichromat passing, it does not ensure that all dichromats fail. In our study, one deuteranope passed both tests (3% of dichromats) regardless of the failure criteria. This subject was 18 yr old female who was diagnosed as deuteranope in her preferred eye and severe deuteranomalous in the other eye. This result was similar to the percentage of dichromats who pass the F-D15 found by Birch (2008) when 1 crossing was allowed to pass (Birch, 2008).

7.6 Conclusion

Without the additional stressor of trying to qualify for a job, this study showed that there was not a strong advantage for requiring a patient to pass on 2/3 versus using the outcome of the 1st trial for either the ColorDx D15 or Farnsworth D15. The subtle advantage of adopting multiple trials within a visit is that it produces marginally better between-session repeatability for the classification of the defect, particularly for the F-D15 because it gives the evaluator more practice in interpreting the patterns. Using a failure criterion of >1 crossing instead of any crossing is recommended because >1 crossing had slightly better

176

between-session repeatability for both pass/fail and classification. The ColorDx D15, which is the computerized version of F-D15, has good agreement with F-D15 and could be an adequate substitute with the caveat that the ColorDx D15 is slightly easier to pass than the F-D15.

Chapter 8

Next Generation of Colour Vision Tests and Predicting the Occupational Tests

8.1 Introduction

In aviation, many countries use different clinical colour vision tests to determine whether individuals have sufficient colour discrimination to qualify for an unrestricted pilot's license. In general, the testing protocol is to run a screening test such as Ishihara first. If the candidate passes the screening, then they are considered to have adequate colour discrimination for an unrestricted license. If the candidate fails the screening test, then another test is carried out to see whether the candidate has a sufficient colour vision to qualify for an unrestricted license. It is the secondary test that varies by country. In Canada, the military and civilian aviation authorities have similar colour vision requirements. In civilian aviation, the pilot is first tested with one of the approved colour vision screening tests. If the candidate fails the screening test, then they must pass either the Farnsworth Munsell D-15 (F-D15) or the Holmes-Wright Lantern Type A (HWA) in order to qualify for an unrestricted pilot's license. The HWA is no longer manufactured, but it still available in different locations within Canada. The Royal Canadian Air Force (RCAF) uses the Ishihara 38 plate edition colour vision test (Ishihara) and the Standard Pseudoisochromatic Plates- Part 2 (SPP2) to screen for colour vision defects. The SPP2 was added primarily to screen for acquired blue-yellow defects. Individuals who fail either screening test are assessed further with the F-D15. If they

178

pass, then they are qualified to become a pilot in the RCAF. In the United Kingdom, candidates for an unrestricted civil aviation license are screened first with Ishihara test. If s/he failed the Ishihara, they will be tested with the Colour Assessment and Diagnosis test (CAD). A number of other countries in the Middle East and Asian are adopting this standard.

The computer-based tests that are currently available, or in development, have several advantages over the traditional printed tests. The first is that several can measure chromatic thresholds for both CVNs and CVDs, which would allow for better monitoring of the pilots' colour vision over their career. Second, they are harder to memorize. Third, there is less administrator bias. Fourth, the test may integrate seamlessly with the electronic medical record. This chapter aims to determine which one of the next generation colour vision tests would be a sufficient substitute for the currently colour vision tests.

8.2 General Methods

The current occupational colour vision tests evaluated in this study are

- Farnsworth D15 test

- Holmes Wright type A (HWA)

- Colour Assessment and Diagnosis (CAD): Pilot criterion

The newer tests used in this chapter are

- ColorDx D15.

- CAD test threshold mode: binocular testing.

- Cambridge Color Vision Test (CCT): binocular testing discrimination ellipse (Area) and Monocular testing (CCT Tri)

- Landolt C cone contrast test (LandC): binocular threshold testing LandC (B), and monocular threshold testing LandC (MT).

- Rabin Cone Contrast Colour Vision test (RCCT): monocular testing.

All the above tests were described previously in Chapter 4. Only the red-green colour-defectives based on the Rayleigh colour match (including the one who was classified as normal by the anomaloscope) were included. The subjects in these studies are described in Chapter 3 section 1.

In evaluating the agreement between the current and newer tests, the AC1 coefficient of the agreement, sensitivity, specificity, and predictive pass/fail were calculated. The predictive pass (PreP) value is the probability that a person who passes one of the newer tests will pass the current test. The predictive fail (PreF) is the probability that a person who fails one of the newer tests will fail the current test.

The parameters evaluated were the ColorDx D15 C-index; CAD red-green threshold; highest of the CCT Tri red-green thresholds and the average of the protan/deutan thresholds, CCT (Area); highest of the L or M cone thresholds and the average of the L and M cone thresholds for both the LandC (MT) and LandC (B); and the lowest of the RCCT L or M-cone sensitivities and the average of the L and M cone sensitivities. The reason for examining both the individual L and M cone thresholds or sensitivities and the average was

to be consistent with CAD red-green SNU. The SNU is the average of the thresholds for the protan and deutan vectors. The monocular data from the CCT Tri, LandC (MT) and RCCT were averaged between eyes. With the exception of the ColorDx D15, Receiver-Operator-Characteristic (ROC) analyses were performed using each test parameter to determine the cut-off score that would give the best agreement with the current occupational test by using the maximum of the sum of the sensitivity and specificity. If there were multiple criteria with the same sum, the cut-off with the highest sensitivity was selected. All data are from the first session.

8.3 Comparison with the F-D15 test

The F-D15 is the current test that determines whether a candidate with a colour vision defect has adequate colour discrimination to qualify for an unrestricted civilian pilots license or RCAF. That is if the candidate fails the test, he or she will not qualify for a position in aviation. The previous chapter compared the F-D15 with the ColorDx D15 using visual inspection. The results showed that although the ColorDx D15 might be slightly easier to pass, there was a high level of agreement. In this chapter, the F-D15 will be compared with the ColorDx D15 using the C-index parameter from the Vingrys and King-Smith analysis in term of pass/fail agreement. Both tests will be evaluated using the suggested C-index failing score of ≥ 1.78. This value corresponds to more than one major crossing.

8.3.1 Results

Table 20 lists the results. The results are arranged in descending order according to the AC1 agreement value. The area under the ROC curve and the AC1 agreement show that the various tests can separate colour-defectives who pass or fail the F-D15 better than chance. Nevertheless, the ColorDx D15 was significantly better than all the rest of tests. The CAD was the next best. The rest of the tests had similar results with moderate levels of agreement.

In general, the PreF values were good-to-excellent on all tests with the ColorDx D15 having perfect and significantly higher PreF than the rest. This means that if the person fails the ColorDx D15, they will almost certainly fail the F-D15. High PreF values indicate that the specificity of the test is very good. However, the reason for the lower AC1 values for the other tests was the decrease in the PreP. The general trend was that PreP was always lower than the PreF. This result indicates that the discrepancy was more likely to be a subject who passed the computer test, but failed the F-D15, than vice versa. For example, the PreP value for the LandC (B) using the highest threshold criterion is 0.74. This means that 26% of the people whose highest threshold is lower than (i.e. better than) -1.048, will still fail the F-D15. A low PreP indicates that the sensitivity of the test relative to the F-D15 is low. The ColorDx D15 PreP was above the rest of the tests and was significantly greater than the LandC (B) average L and M cone threshold, CCT (Area) and the monocular test values, but not statistically different from the CAD and maximum threshold for the LandC (B) PreP values.

182

Table 20. Results of the ROC and agreement analyses for the comparison between the F-D15 and selected tests.

Test	ROC Area (95% CI)	Cut-off point	AC1 (95% CI)	PreP (95% CI)	PreF (95% CI)
ColorDx D15	NA	1.78	0.88 (0.77 to 0.99)	0.87 (0.71 to 0.95)	1 (0.9 to 1)
CAD	0.89 (0.81 to 0.97)	20.05	0.71 (0.54 to 0.88)	0.82 (0.63 to 0.92)	0.87 (0.72 to 0.94)
LandC (B) Maximum Threshold	0.75 (0.63 to 0.87)	-1.048	0.53 (0.32 to 0.74)	0.74 (0.52 to 0.87)	0.76 (0.6 to 0.85)
RCCT Minimum Sensitivity	0.73 (0.61 to 0.86)	41.25	0.51 (0.3 to 0.72)	0.69 (0.52 to 0.84)	0.79 (0.64 to 0.89)
CCT Tri Maximum Threshold	0.76 (0.64 to 0.89)	532	0.51 (0.31 to 0.73)	0.70 (0.5 to 0.83)	0.78 (0.62 to 0.87)
CCT Tri Avg	0.76 (0.64 to 0.89)	474.8	0.47 (0.26 to 0.69)	0.65 (0.48 to 0.78)	0.82 (0.66 to 0.92)
RCCT Average Sensitivity of L and M-cones	0.69 (0.55 to 0.82)	66.88	0.46 (0.24 to 0.68)	0.67 (0.48 to 0.81)	0.76 (0.61 to 0.86)
LandC (B) Average of L and M-cone Thresholds	0.76 (0.65 to 0.88)	-1.491	0.45 (0.23 to 0.67)	0.65 (0.47 to 0.79)	0.78 (0.63 to 0.89)
LandC (MT) Maximum Threshold	0.78 (0.67 to 0.89)	-1.088	0.44 (0.23 to 0.67)	0.63 (0.44 to 0.82)	0.82 (0.42 to 0.96)
CCT (Area)	0.72 (0.59 to 0.85)	$1.1*10^{-3}$	0.44 (0.22 to 0.66)	0.62 (0.53 to 0.87)	0.84 (0.61 to 0.86)
LandC (MT) Average of L and M-cone Thresholds	0.72 (0.59 to 0.84)	-1.292	0.3 (0.065 to 0.53)	0.55 (0.4 to 0.68)	0.83 (0.64 to 0.93)

8.4 Comparison with the Holmes-Wright Type A (HWA)

The HWA lantern test determines whether a person has the adequate colour vision to identify small lights used in aviation and the maritime industry in Canada.

8.4.1 Results

Table 21 lists the results of the ROC analysis, cut-off points, AC1 values, PreP and PreF, and 95% confidence intervals for the various tests. The F-D15 is included for reference. Despite the small number of colour-defectives who passed the HWA, the area under the ROC curves was significantly greater than chance and the level of agreement was statistically identical to 1.0 in several cases. The notable exception was the F-D15, which had a moderate level of agreement with the HWA. The PreP was equal to 1.0 for several tests although the precision of the value is low because of the small number of a subject who passed the HWA. Because of the low statistical power, only the PreP values for the maximum threshold of the CCT Tri and F-D15 were significantly lower.

Table 21. Results of the ROC and agreement analyses for the comparison between the HWA and selected tests.

Test	ROC Area (95% CI)	Cut-off point	AC1 (95% CI)	PreP (95% CI)	PreF (95% CI)
LandC (MT) Average of L and M-cone Thresholds	0.92 (0.72 to 1.08)	-1.45	0.98 (0.95 to 1.00)	1.00 (0.51 to 1)	0.98 (0.92 to 1.00)
LandC (MT) Maximum Threshold	0.90 (0.78 to 1.06)	-1.62	0.98 (0.95 to 1.00)	1.00 (0.51 to 1)	0.98 (0.92 to 1.00)
LandC (B) Average of L and M-cone Thresholds	0.93 (0.80 to 1.06	-1.78	0.98 (0.95 to 1.00)	1.00 (0.51 to 1)	0.98 (0.92 to 1.00)
LandC (B) Maximum Threshold	0.96 (0.88 to 1.04)	-1.57	0.98 (0.95 to 1.00)	1.00 (0.51 to 1)	0.98 (0.92 to 1.00)
CCT Tri Average of L and M-cone Thresholds	0.94 (0.83 to 1.06)	122.5	0.98 (0.95 to 1.00)	1.00 (0.51 to 1)	0.98 (0.92 to1.00)
CAD	0.87 (0.65 to 1.10)	3.12	0.97 (0.92 to 1.00)	0.80 (0.38 to 0.96)	0.98 (0.92 to 1.00)
RCCT Average Sensitivity of L and M-cones	0.82 (0.48 to 1.16)	81.9	0.95 (0.89 to 1.00)	0.67 (0.30 to 0.90)	0.98 (0.91 to 1.00)
RCCT Minimum Sensitivity	0.89 (0.69 to 1.09)	71.3	0.95 (0.89 to 1.00)	0.67 (0.30 to 0.90)	0.98 (0.91 to 1.00)
CCT (Area)	0.98 (0.94 to 1.01)	$1.95*10^{-3}$	0.93 (0.85 to 1.00)	0.56 (0.27 to 0.81)	1.00 (0.94 to 1.00)
CCT Tri Maximum Threshold	0.97 (0.90 to 1.03)	404.5	0.80 (0.69 to 0.94)	0.33 (0.15 to 0.58)	1.00 (0.93 to 1.00)
F-D15	NA	NA	0.42 (0.18 to 0.65)	0.14 (0.057 to 0.31)	0.98 (0.87 to 1.00)

8.5 Comparison with the CAD Pilot

In the previous sections, we determined the level of agreement between the CAD and other tests by treating the cut-off value of the CAD as a free parameter. However, a number of civilian aviation authorities have adopted the CAD test to determine whether pilots have sufficient colour discrimination to qualify for an unrestricted commercial pilot's license. The failure criteria for an unrestricted license are a red-green threshold greater than 6 SNUs for deutans and greater than 12 SNUs for protans (Civil Aviation Authority, 2009).

8.5.1 Results

Similar to the HWA, only a few subjects, one protan, and seven deutans, met the CAD pilot criterion. Because only one protan subject passed the CAD pilot criterion, no ROC analysis was performed on this group. Because the CAD pilot criterion depends on the type of red-green defect, the analysis of the deutan results included any subject whose defect was unclassified by a given test or had a deutan defect and was classified as CVN by one of the newer computer tests.

Although none of the colour-normals and deutans were misclassified as protan on CAD, LandC (MT), LandC (B) and RCCT, there were many discrepancies with the CCT Tri. Twenty-one percent (n=8 deutans and n=6 protans) had different classifications for each eye on the CCT Tri. In addition, the CCT Tri misclassified 46% of the protan subjects as deutans in both eyes. Although none of the subjects in this last group had red-green thresholds below 12 SNU on the CAD, it does raise concerns about the ability of the CCT Tri to classify the nature of the red-green defect when that information is important. Also from the individuals

186

who failed the CCT Tri only 7.5% of the deutan (n=3) meet the CAD criterion for deutan. Because of a large number of misclassifications of the protans and a small number of deutan who meet the CAD pilot criterion, the CCT Tri was not evaluated further.

Table 22 lists results of the ROC and agreement analyses. The LandC (B) had two cut-off points with almost the same maximum sum of sensitivity and specificity (1.35 vs 1.34). Using the larger sum (1.35) was resulted in a chance agreement with the CAD test. However, the other one (1.34) had a much higher level of agreement and correlated with the monocular cut-off value. For these reasons, we selected latter cut-off for further analysis.

Only the LandC (MT) and LandC (B) maximum threshold had an area under the ROC curve that was significantly above 0.5. Nevertheless, if the area under the curve was greater than 0.50 but not significantly different from chance, the AC1 level of the agreement values were calculated because the statistical power might be different between the procedures. If the AC1 value was not significantly different from zero, then no further calculations were performed.

Although the agreement between the CAD-pilot, RCCT, and some of the LandC tests parameters were significantly better than zero, the level of agreement was only moderate. The PreF values of around 0.85 were good, but it is still possible that 15% of the people who fail the LandC or RCCT would pass the CAD-pilot criterion. The PreP values, however, were generally lower. Except for the average of the LandC L and M-cone thresholds for monocular and binocular viewing, which had PreP values of at least 0.66, no more than 50% of the people who passed the other tests would pass the CAD-pilot criterion. Even for the

LandC average monocular thresholds, there were still 25% of the individuals who passed this

test but failed the CAD pilot. Again, because so few individuals met the CAD pilot criterion,

the statistical power for comparing the PreP was very low and none of the values was

significantly different.

Table 22. Results of the ROC and agreement analyses for the comparison between the CAD pilot criterion and the selected tests.

Test	ROC Area (95% CI)	Cut-off point	AC1 (95% CI)	PreP (95% CI)	PreF (95% CI)
LandC (MT) Average of L and M-cone Thresholds	0.68 (0.46 To 0.90)	-1.62	0.80 (0.63 to 0.97)	0.75 (0.30 to 0.95)	0.86 (0.71 to 0.94)
LandC (B) Average of L and M-cone Thresholds	0.67 (0.46 To 0.88)	-1.78	0.77 (0.58 to 0.96)	0.66 (0.2 to 0.93)	0.84 (0.68 to 0.92)
RCCT Average Sensitivity of L and M-cones	0.51 (0.22 To 0.80)	81.9	0.72 (0.51 to 0.93)	0.5 (0.19 to 0.81)	0.85 (0.70 to 0.94)
RCCT Minimum Sensitivity	0.57 (0.32 To 0.82)	71.3	0.72 (0.51 to 0.93)	0.5 (0.19 to 0.81)	0.85 (0.70 to 0.94)
LandC (B) Maximum Threshold	0.75 (0.57 To 0.93)	-1.28	0.52 (0.23 to 0.80)	0.41 (0.22 to 0.64)	0.96 (0.79 to 0.99)
LandC (MT) Maximum Threshold	0.76 (0.58 To 0.95)	-1.11	0.36 (0.042 to 0.67)	0.35 (0.18 to 0.57)	0.95 (0.76 to 0.99)

8.6 Discussion

8.6.1 F-D15 test

The current RCAF will accept individuals who have a colour vision defect and pass the F-

D15. If one of the newer tests replaces the F-D15, then the new test should have a very good

agreement with the F-D15 and high predictive values for passing or failing the F-D15. Except for the ColorDx D15, the agreement with the other computer-based tests was only moderate. The PreF values ranged from 0.76 to 0.87. This means that 13% to 24% of the candidates who fail one of the newer computer tests could pass the F-D15. A low PreF means that the tests have a low specificity (more false positives). The PreP values are marginally lower, ranging from 0.55 to 0.82. That is, between 18% and 45% of the candidates who pass one of the newer computer tests could fail the F-D15. Low PreP means that the computer tests have a low sensitivity (more false negatives) relative to the standard test. Of the newer computer threshold tests, the CAD would be the better option because its PreP is 0.82 and the PreF is 0.87, but because the values are not perfect, 31% of the CVD taking the CAD would be misclassified.

In terms of comparing the F-D15 with the ColorDx D15, the agreement between the two tests using visual inspection, which has been determined in chapter 7, was relatively high. In this chapter, however, we compared the two tests using the C-index from Vingrys and King-Smith parameters. This was done for two reasons. First, there will be no examiner bias in interpreting the results when using the C-index. Second, the C-index takes into account both crossings and transpositions. This would help in interpreting results where there is one crossing and multiple transpositions, although none of the subjects in this study had one crossing and multiple transpositions. The ColorDx D15 had excellent agreement with the F-D15 using the Vingrys and King-Smith parameter. This level of agreements was the same when using the visual inspection with >1 crossing as a failure criterion. The PreF value using

the C-index was 100%, and it was significantly greater than the rest of the binocular and monocular test values. The slightly lower PreP indicates that the ColorDx D15 is slightly less sensitive than the F-D15. One possible reason for the lower sensitivity is that the circles on the ColorDx D15 are twice the size of the F-D15. Colours are slightly easier to identify for larger objects (Cole et al., 2006b).

8.6.2 HWA test

The HWA determines whether individuals with a colour vision defect can identify aviation signal lights that are viewed from 1 nautical mile (Holmes and Wright, 1982; Cole and Vingrys, 1982). Although still in limited use, the HWA is no longer manufactured and could be considered as a legacy test. Nevertheless, the HWA does have face validity with the colour and intensity of the maritime and aviation running lights, and so the results from this study do show how well the individual tests predict performance in identifying these signal lights. The low pass rate for the CVD group is consistent with previous studies showing that only individuals with mild defects have the highest probability of passing, especially if multiple runs of the nine pairs of lights are always presented (Vingrys and Cole, 1983; Birch, 2008b; Hovis, 2008). The fact that the cut-off value for the LandC tests was only 0.20 log units higher than the recommended pass/fail and the cut-off score for the RCCT was very close to the recommended pass/fail score of 75. Also supports the conclusion that only individuals with very mild or mild colour vision defects will be able to pass the HWA. Changing the RCCT cut-off to 75 affected the outcome of 1 deuteranomalous to fail both

tests and another deuteranomalous to fail the RCCT but pass the HWA. This resulted in little change in the agreement and predictive values. The result that only the individuals with mild defect who either passed the computer-based or had a threshold that was only slightly evaluated explains the near perfect PreF. Essentially, if the candidate is classified as CVD on one of the newer tests based on the manufacturer's scoring criterion, then there is over a 90% chance that they will fail the HWA. This agrees with Cole and Vingrys' (1983) results.

The low PreP for the F-D15 was also in agreement with Cole and Vingrys' (1983) previous study. The F-D15 is designed to separate those individuals with a moderate-to-severe defect from those with a milder defect. However, the HWA fails a large percentage of the mild-to-moderate cases, and so the F-D15 PreP should be relatively low. The ColorDx D-15 would not offer any improvement because it was marginally less sensitive than the F-D15.

Although the PreP values for the LandC tests were very good, the uncertainty was relatively large. This was due to a small number of subjects who passed the HWA. This small number could be a result of our recruiting process. It is possible that individuals with a mild defect were either unaware of their colour vision defect or were not interested in participating in the study. Nevertheless, the 7% pass rate in our study was within the 7% to 15% rate reported by others who used results from clinical files (Vingrys, 1983; Birch, 2008b). The range in the pass rate was partially due to the scoring procedure. The standard scoring criterion results in a higher pass rate because a number of CVD can pass based on their performance on the first presentation of the nine pairs (Hovis, 2008; Birch, 2008b). If the pass/fail criterion is adopted that is based onmultiple presentations of the nine pairs of lights

191

within a session, then the pass rate drops to less than 10% of the CVDs, but the between-visits repeatability improves (Hovis, 2008; Birch, 2008b).

8.6.3 CAD-pilot criterion test

The CAD pilot criterion is based on a task analysis showing that identifying Precision Approach Path Indicator (PAPI) lights is a colour-critical task in civilian aviation (Civil Aviation Authority, 2006b, 2009). The PAPI lights are also used in military aviation, but the on-going task analysis may show that there is a different colour-critical task. Nevertheless, a number of civil authorities use the CAD pilot criterion, and so it is useful to compare the CAD with the other computer tests to determine whether they could be adequate substitutes. Although, the CAD-pilot criterion is actually less stringent than the pass/fail criterion used to determine normal versus red-green defective colour vision, only 12% of the CVD group met this requirement. Again, this could have resulted from the recruiting process. Nevertheless, the agreement between the CAD pilot criterion and the other tests was only moderate. Although the LandC (MT) average of the L and M-cone, RCCT average sensitivities of L and M cones, and the LandC (B) maximum and average threshold had a reasonable PreF for failing the CAD-pilot criterion, the PreP was only 0.50 for most of these tests. This value indicates that 50% of the deutan colour-defectives who pass the other computer based colour vision tests will not meet the CAD criterion. The exception was the LandC (MT) average of the L and M cone thresholds, which had PreF of 0.75; however, this value was not

192

statistically different from the other values because of the small number of subjects who met the pilot criterion.

8.7 Conclusion

In terms of selecting an alternative colour vision test battery, the ColorDx D15 test would be the best option to replace the F-D15. The comparison with the HWA confirmed that only CVD with a mild colour defect can pass this test and so the newer computer tests had a reasonable level of agreement and high predictive values for failure because approximately 90% of the CVDs will fail the HWA. In comparison with CAD pilot colour vision test, no one test stands out as being superior to the rest. The low agreement with other test might be due to the small number of subjects in our study who meet CAD pilot criterion.

Chapter 9

The CN Lantern Test and different viewing distances

9.1 Introduction

Trains in many countries are controlled by the block system, which divides the rail network into intervals of varying length (Hovis and Oliphant, 1998). Coloured signal lights are used to grant permission to enter the block and set the speed limit. In Canada, the coloured lights are red, green and yellow. The lights may be presented as a single signal light, a pair of signal lights or a triplet of signal lights (Hovis and Oliphant, 1998). The coloured light signals on the main track are usually referred to wayside signals. Wayside signals are typically 5 m above the track which puts them about 1m above the eye level of the locomotive engineer (Hovis and Ramaswamy, 2006). Sighting distances for railway signal lights vary depending on whether the train is operated in the yard or on the main track. On the main track, the sighing distance for the wayside signals can vary from 300 to 1600 m. From these viewing distances, the angular subtense of light ranges from 0.5 to 2 arc-min so that they essentially point sources (Hovis and Ramaswamy, 2006). In the yard, the sighting distances can be within 100 m.

Identifying wayside signal lights from a long sighting distance is challenging for individuals with a colour vision defect (CVD) (Hovis and Oliphant, 1998, 2000; Dain et al., 2015). However, individuals with a CVD who do not qualify for positions requiring longer sighting distances may qualify for positions using shorter sighting distances. Individuals with

194

a CVD make fewer errors when the coloured object gets larger or brighter (Kinney et al., 1979; Vingrys and Cole, 1983; Steward and Cole, 1989; Hovis and Ramaswamy, 2006).

One of the purposes of this study was to extend the Hovis and Ramaswamy study (2006) which looked at the CN Lantern pass rate at 4.6m and 2.3 to shorter viewing distances. These two distances would correspond to a viewing distance range of 200 to 650m and 100 to 330m respectively. Changing the viewing distance from 4.6 m to 2.3 m increased the pass rate for CVDs from 8% to 22%. We wanted to determine the rates for 1.15m and 0.57m. These two shorter test distances would correspond to a range of 50-163m and 25-82m on the track respectively. The second purpose was to determine the repeatability of the lantern results in terms of pass/fail at these distances. The third purpose of the study was to determine whether various clinical colour vision tests could predict the lantern test outcome at the various distances. These later results will be presented in the next chapter.

9.2 Subjects

As mentioned in chapter 3 section 2, the railway study recruited Fifty-six CVN subjects (CVN) and 62 subjects with a congenital red-green CVD. More details of the participants in this chapter were presented in chapter 3 section 2.

9.3 Methods

The CN Lantern is described in detail in section 4.2; however, the following is a brief summary for the reader. The CN Lantern presents 15 different sets of three lights. The lights are displayed one above the other, and each is surrounded by a small black annulus. The first

two presentations are for demonstration, and the next 13 are the test lights. The subject's task

is to name the colours from top to bottom. Each light in the triplet is scored as a correct or

incorrect response. The order of test distances starts at the longest distance and progressively

decreases as if the train is moving toward the signal light. Providing the error was not naming

red as green or vice versa, one error was allowed at the longest distance of 4.6 m. No errors

were allowed at the other test distances (Hovis and Ramaswamy, 2006).

The pseudoisochromatic plate tests (Ishihara, HRR, and ColorDx PIP) evaluated in

this study were administrated in between the CN Lantern test runs. This was done to reduce

the practice effect on the CN Lantern. The order of the test light presentation was the

opposite of the order from the previous test distance starting from the very last presentation.

Participants were asked to return after 10 to 15 days for the second round of testing

9.4 . Results

9.4.1 Viewing distance influences the error rate

All the CVN subjects passed the CN lantern at both visits without any errors. Figure 52

shows the average number of errors made by each type of defect at the first session. On

average, the number of errors made by anomalous trichromats was less than made by

dichromats. Protans made fewer errors than deutans. As expected, the number of errors

decreased with decreasing the viewing distance. Repeated measures analysis of variance

showed that the effect of distance ($F_{(2.09)} = 97.469$; p<0.005) and type of defect were

significant ($F_{(3)} = 6.511$; p=0.001). The interaction term was not significant ($F_{(6.3)} = 0.836$;

p=0.549). The Greenhouse-Geisser adjustment was used for the within-subjects effects and interaction because the data did not meet the sphericity assumption. Post Hoc testing (Dunnett T3 for unequal variances) indicated that the deuteranopes made significantly more errors than the other 3 groups (p<0.027). None of the other comparisons between the different colour-defect groups was significant.
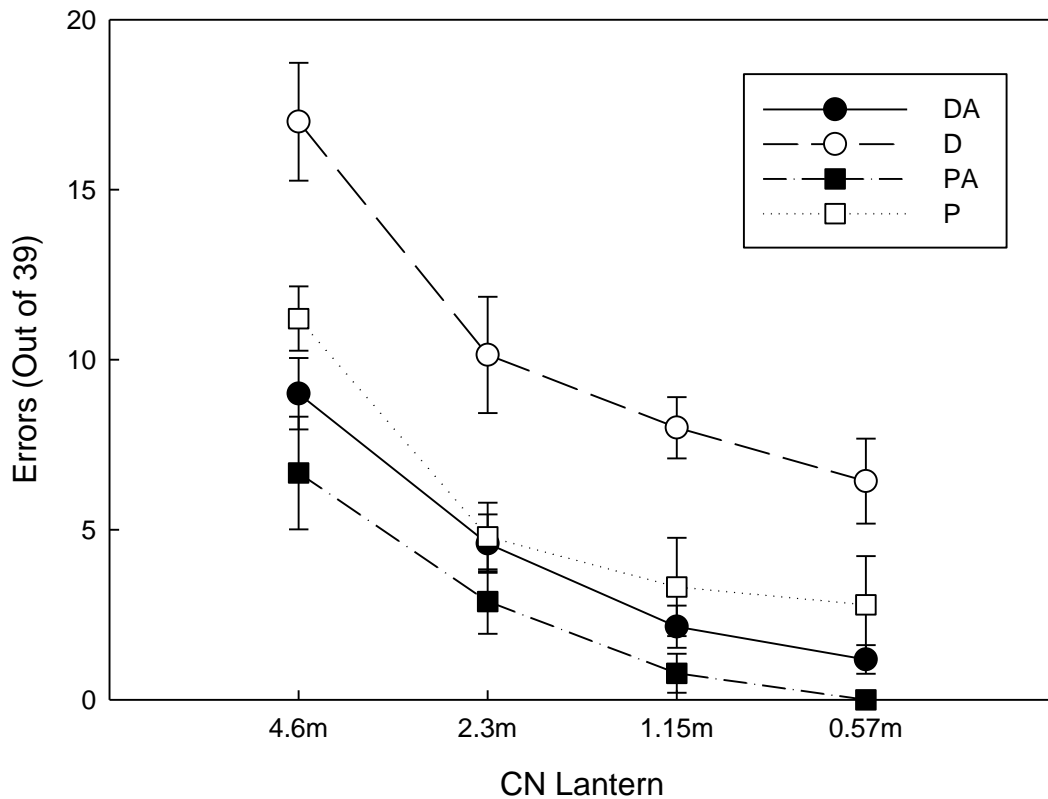
Figure 52. Mean of number of errors on the CN Lantern test by each type of defect at each distance. DA is deuteranomalous, PA is protanomalous, D is deuteranope and P is protanomalous. The error bars are the standard error.

The reduction in the number of errors with decreasing viewing distance translated into increasing the overall proportion of participants who passed the CN lantern. Figure 53 shows the percentage of participants that passed for each type of defect at each distance at the first session. All participants who passed at a longer distance also passed at the shorter ones. The anomalous trichromats had a high passing rate with shorter distances. However, 26% and 42% of the protanopes passed at 1.15m and 0.57m respectively. Only one deuteranope subject passed at the shortest distance (0.57m).
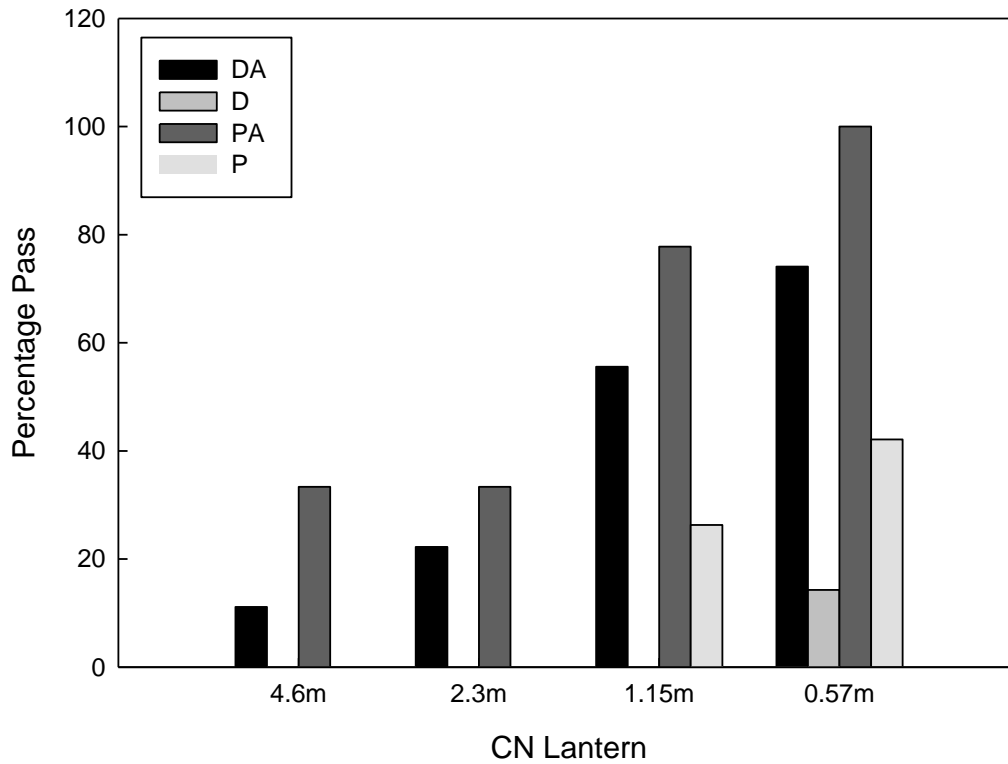
Figure 53. Percentage of each type of defect who passed the CN lantern test at each viewing distance at the first session. DA is deuteranomalous, PA is protanomalous, D is deuteranope and P is protanomalous.

9.4.2 Repeatability of CN lantern

Table 23, 24, 25 and 26 show the repeatability of the CN lantern test between first and second session for the CVDs at 4.6m, 2.3m, 1.15 m and 0.57m respectively. In general, the AC1 coefficients of agreement for between session pass/fail were very good. The one exception was the coefficient at 0.57 m, which was moderate and significantly lower than the AC1 at all other distances based on the 95% Confidence Intervals. Sixty-five percent of the discrepancies across all distances are those who failed the first session but passed the second while approximately 35% passed the first and failed the second session.

Cicchette and Feinstein (1990b) recommended calculating the agreement for passing (P_{agree}) and agreement for failing (F_{agree}) separately when there is a large asymmetry between the marginal totals in a comparison table (Cicchette and Feinstein, 1990) as is the case in Tables 23 and 24. These two indices represent the number of people who pass (or fail) both sessions divided by the average number of passing or failing at either session. The F_{agree} values between the first and second sessions at most distances were greater than 0.94. The exception was F_{agree} of 0.83 at the shortest distance. It was significantly lower than the F_{agree} values at 4.6m and 2.3m based on the 95% confidence interval. The P_{agree} values were 0.67 at the two longer distances and approximately 0.90 for the shorter distances. The large confidence intervals at the longer distances were a result of the small number of subjects who passed at either session.

Table 23. The repeatability between the first and second session of the CN lantern at 4.6m

First Session		
4.6m	**Pass**	**Fail**
Pass	3	2
Fail	1	52

Second Session (row labels for Pass/Fail)

P_{agree} = 0.67 (0.30 to 0.903)

F_{agree} = 0.97 (0.89 to 0.99)

AC1= 0.94, 95% Confidence Interval: 0.87 to 1

Table 24. The repeatability between the first and second session of the CN lantern at 2.3m

First Session		
2.3m	**Pass**	**Fail**
Pass	6	5
Fail	1	46

P_{agree} = 0.67 (0.39 to 0.86)

F_{agree} = 0.94 (0.84 to 0.98)

AC1= 0.86, 95% Confidence Interval: 0.74 to 0.98

Table 25. The repeatability between the first and second session of the CN lantern at 1.15m

First Session		
1.15m	**Pass**	**Fail**
Pass	24	1
Fail	1	32

P_{agree} = 0.96 (0.81 to 0.99)

F_{agree} = 0.97 (0.85 to 0.99)

AC1= 0.93, 95% Confidence Interval: 0.84 to 1

Table 26. The repeatability between the first and second session of the CN lantern at 0.57m

First Session		
0.57m	**Pass**	**Fail**
Pass	31	5
Fail	3	19

P_{agree} = 0.89 (0.75 to 0.95)

F_{agree} = 0.83 (0.65 to 0.93)

AC1= 0.74, 95% Confidence Interval: 0.57 to 0.91

Table 27 shows the CVD subjects (dichromats and anomalous trichromats) who had discrepancies in their pass/fail outcomes at the two sessions. Twenty-two percent of the anomalous trichromats (6 deuteranomalous and 3 protanomalous) had discrepancies between visits compared to 30% of the dichromats (7 protanopes and 1 deuteranope). Only one deuteranomalous subject and one protanope showed two discrepancies at two different distances. The deuteranomalous (21yr old) individual passed the CN lantern at 2.3m and 1.15 m at the first visit but failed at both distances at the second visit with one error at each distance. The protanope (21 yr old) failed at 2.3m and 0.57m in the first session with 2 errors and passed at both distances at the second session.

Table 27. The discrepancies between the first and second sessions for the CN lantern at all distances. FP indicates passing the first session and failing the second session and PF passing the first session and failing the second.

Diagnosis	4.6m		2.3m		1.15m		0.57m	
	FP	PF	FP	PF	FP	PF	FP	PF
Deuteranomalous			✓					
Protanomalous			✓					
Protanomalous			✓					
Protanope							✓	
Deuteranomalous				✓		✓		
Protanope			✓				✓	
Protanope							✓	
Protanope							✓	
Deuteranomalous	✓							
Protanope			✓					
Protanomalous		✓						
Protanope					✓			
Deuteranomalous								✓
Deuteranope								✓
Deuteranomalous	✓							
Protanope							✓	
Deuteranomalous								✓

The finding that more subjects passed the CN lantern at the second session means that the number of errors made at the second session should be lower. Repeated measures analysis of variance with the Greenhouse-Geisser correction for sphericity showed that the effect of session ($F_{(1)} = 5.863$; $p<0.019$) and distance were significant ($F_{(1.78)} = 121.595$; $p<0.001$). The interactions term between distance and session, session and diagnosis, distance and diagnosis, distance and session and diagnosis were not significant ($p>0.258$). This indicates that all subjects were performing the first and second session in the same

manner across the four distance, but had fewer errors at the second visit. Figure 54 shows

the average of the difference of errors between the first and second visit (subtracting the 2nd

visit from the 1st visit). Positive values indicate that there are more errors at the first visit. In

general, there were more errors at the first visit for all types of colour defect practically for

the longest distance. Only the deuteranopes at 2.3m had an average difference in the negative

section indicating that there were more errors at the second session at this particular distance.
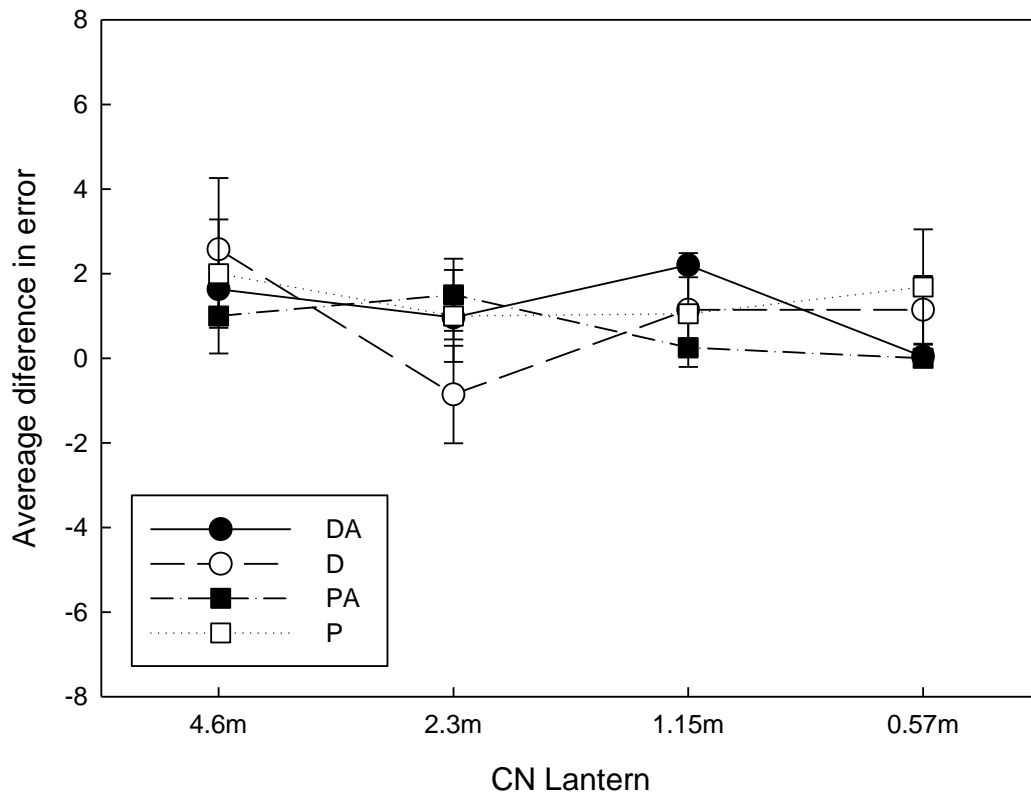
Figure 54. The average error differences between the first and second visit. Positive values indicate more errors at the first visit. DA is deuteranomalous, PA is protanomalous, D is deuteranope and P is protanomalous. The error bars are the standard error.

9.5 Discussion

The reduction in the number of errors with decreasing the viewing distance and corresponding increase in pass rate is consistent with the fact that the individuals with colour vision deficiency make fewer errors as the test lights get brighter (Kinney et al., 1979; Vingrys & Cole, 1983; Steward & Cole, 1989; Hovis & Ramaswamy, 2006). Increasing the retinal illumination was the primary reason for making fewer errors because the lights are essentially point sources at all distances. Each time the viewing distance was decreased by a factor of 2, the retinal illumination increased by a factor of 4 so that the retinal illumination of the lights at the closest viewing distance was 64 times brighter than at the farthest distance. The largest decrease in the average number of errors was by 50% when moving from 4.6m to 2.3 m. The smallest decrease in the average number of errors was 30% when moving from 1.15 m to 0.57m. This suggests that the increasing retinal illumination of point sources higher than the values at the 0.57 m distance offers little advantage to the colour-defectives. Based on the lantern's design (Hovis & Oliphant, 2000), this would be approximately 2.3 log units above the colour normal threshold.

As expected, none of the dichromats passed the CN lantern at 4.6m in this study. This is similar to Hovis and Ramaswamy's (2006) findings that dichromats rarely pass the CN Lantern at the longest viewing distance. In general, few CVD subjects passed the CN Lantern at the longest distance and ones that did they were anomalous trichromats (Hovis & Ramaswamy, 2006; Hovis and Oliphant, 2000). Decreasing the viewing distance allowed additional anomalous trichromats to pass (10% of the deuteranomalous and 35 % of the

protanomalous), which is similar to what Hovis and Ramaswamy (2006) found in their study

(5% of the deuteranomalous and 20% of the protanomalous). At the two shortest distances,

the pass rate continued to improve with all the protanomalous, approximately 75% of

deuteranomalous and a few dichromats passing at the 0.57m.

When the results were analysed by whether the defect was classified as protan versus

deutan, the percentage of deutans who passed the test was larger than the percentage of

protans at each distance. This would suggest that protans have more difficulty in recognizing

signal lights. However, there was a brightness difference added to some tests to confuse

protans, and that could also be the reason for the smaller percentage of protans who pass the

CN lantern. The result that 59% of the protans passed the lantern at the shorter distances

suggests that this brightness differences may not be sufficient when the overall brightness is

increased (Hovis and Oliphant, 2000).

Although there was a decrease in the number of errors at the second session, the

repeatability of CN lantern on a pass/fail basis was very good at most distances. The

exception was at 0.57m (AC1= 0.74) which was significantly lower than the agreement at

4.6m. Because there was an overall reduction in the number of errors, likely due to a learning

effect, most of the discrepancies were an improvement to a pass at the other distances.

Nevertheless, we are uncertain as to why all the improvements to pass were protanopes. With

the exception of two subjects who passed at 2.3m and one at 1.15, none of other protanopes

passed at any of the longer distances. This suggests that there may have been secondary clues

present at the 0.57 m that these protanopes were able to use to their advantage.

Identifying individuals who would likely pass the test at the second session would desirable from the candidate's perspective if they failed the Lantern test; however, identifying the individuals who passed the first session, but failed the second session is, equally, if not more, important from the safety perspective. These individuals were one protanomalous at 4.6m, one deuteranomalous at 2.3, 1.15 m and 2 deuteranomalous and one deuteranope at the shortest distance. All of these individuals passed both F-D15 and ColorDx D15 at both visits. The only exception was one of the deuteranomalous at the shortest distance.

Although there was an overall reduction in the number of errors by 30% acrossod all conditions at the second session, the result that none of the interactions that involved both distance and session were significant suggests that the change in the errors as function of test distance was similar in both sessions and further supports the hypothesis that in the increase in in retinal illumination with the decreasing viewing was responsible for the reduction in errors with viewing distance.

9.6 Conclusion

The study showed that the number of errors is decreased when the retinal illumination increased by decreasing the viewing distance. Most of this reduction in the number of errors was within the anomalous trichromat group. However, there was a small number of dichromats passed the test at the shorter distances. The CN lantern test on a pass/fail basis was a repeatable test at all distances even though the average number of errors decreased at

the second session. The study showed that more than 65% of the CVDs could work on the

railway yard safely if the sighting distances were within 82 m.

Chapter 10

Predicting the CN Lantern Test for Railways with clinical colour vision tests

10.1 Introduction

Colour vision requirements for locomotive engineers and conductors vary according to jurisdiction. In the United States, the Federal Railroad Administration (FRA) states that locomotive engineers and conductors must have the "ability to recognize and distinguish between the colours of railroad signals" (Federal Railroad Administration, 2015). This is usually interpreted to mean that the person must pass one of the recommended colour vision tests, although individual companies may develop and use a practical field test (Federal Railroad Administration, 2015). Several railroad companies in Canada and Australia require individuals with a CVD to pass a lantern test which either has face validity with the task (Dain et al., 2015) or has been validated against wayside signal displays (Hovis and Oliphant, 1998, 2000)

One of the factors that can affect the performance of CVDs in identifying signal lights is the brightness of the light (Cole and Brown, 1966; Nathan et al., 1964; Kinney et al. 1979; and Paramei, 1998). For point sources, brightness is often referred to as point brilliance, which is the illumination of the signal light at the corneal plane. On the railway, this translates to the sighting distance at which the engineer and conductor can first identify the

colour of the light. The inverse square law states that the point brilliance of the signal light increases by the square of the reduction in viewing distance.

The point brilliance of CN Lantern test lights are scaled to be equivalent to sighting distances of 0.2–0.65 km at a 4.6 m viewing distance (Hovis and Oliphant, 2000). Previous work has shown that the pass rate increases when the test distance is reduced to 2.3m, which would be equivalent to a range of 0.1-0.33 km in the field (Hovis and Ramaswamy, 2006). It is possible that the sighting distances could be even shorter in the rail yard or if the locomotive was travelling at a slow speed. For this reason, we extended the range of test distances for the CN lantern. The purpose of this chapter is to determine how well various clinical colour vision tests could predict CN lantern results at various test distances. Because the CN lantern is administered by the railway companies or contractors, it is limited in availability and so the information from this study could help clinicians in counselling their CVD patients who may be interested in a career as a locomotive engineer or conductor.

10.2 Method

The dimensions of the CN lantern test have been described in detail in Chapters 4 and 9. The clinical colour vision tests used in this study were HRR, Ishihara, ColorDx PIP, RCCT, CAD, CCT, LandC, F-D15, and ColorDx D15. All the instructions of the tests were presented previously in chapter 4. With the exception of the RCCT, the tests were performed binocularly. The average of the RCCT right and left were used in this analysis. The Ishihara test was included because the railroad companies use this test to determine who has a red-

green CVD and needs to be further tested with the CN lantern. One pseudoisochromatic colour vision (Ishihara, HRR, and ColorDx PIP) tests were administrated to the subject in between each viewing distance on the CN Lantern test. However, the other tests were done on a different day. These tests were RCCT, CAD, CCT, LandC, Farnsworth D15 and ColorDx D15.

10.3 Subjects

Subjects recruited in this study were the same individuals described in Chapter 3 section 2.

10.4 Analysis

The agreement between the CN lantern and the other colour vision tests will depend on which test parameter is used. The parameters evaluated were the qualitative severity of the defect for the HRR and ColorDx PIP, the lowest of the RCCT L or M-cone sensitivities and the average of the L and M cone sensitivities, CAD red-green threshold, CCT (Area and Ellipse Length), highest of the LandC L or M cone thresholds and the average of the L and M cone thresholds, and average number of crossings and C-index for three trials of the Farnsworth D15 and ColorDx D15.

In order to determine the appropriate cut-off value, we used Receiver-Operator-Characteristic (ROC) analyses (Sigmaplot ver 11.0; Systat Software Inc, Chicago, IL). If the area under the ROC curve was significantly greater than 0.5 based on the 95% confidence interval, then the optimal cut-off value was determined. The optimum cut-off score was based on the maximum sum of the sensitivity and specificity. If there were multiple criteria

with the same maximum value, then the cut-off score was selected based on which one of the choices had the highest sensitivity. Once the cut-off value was found, the AC1 coefficient of agreement and predictive pass/fail were calculated. The predictive pass (PreP) value is the probability that a person who passes a clinical test also passes the CN lantern. The predictive fail (PreF) is the probability that a person who fails a clinical test fails the CN lantern.

10.5 Results

Although the anomaloscope was used to diagnose the type of colour vision defect, the Ishihara results were used to classify the subjects as CVN or CVD. The reason for using this test to classify the colour vision was that only the candidates who fail the Ishihara test have to take the lantern test. All the CVNs and three-CVDs (two deuteranomalous and one protanomalous) passed the Ishihara test. All of these subjects passed the CN lantern at all distances without error. They also passed all the other clinical tests in this study.

Table A in Appendix 4 lists the ROC area, the cut-off point, AC1, PreP and PreF values. There were a number of test conditions where the area under the ROC curve was not statistically different from 0.50. This included the RCCT and LandC tests at all CN lantern test distances; HRR, ColorDx PIP, and CAD at 4.6m; and CCT area and ellipse length at 4.6m, 2.3m, and 1.15m. Table 28 lists the cut-off values for each test in which the ROC curve area was significantly greater than 0.5.

Figure 55 shows the AC1 coefficient agreement between the remaining tests and the CN lantern at the various test distances. Note that the AC1 coefficients of agreement between

the F-D15 and CN lantern tests using a number of crossings was slightly different than the values calculated using the C-index at each distance. The differences were not significant based on the 95% confidence intervals. The AC1 value of the C-index was selected over the number of crossings for this analysis because it was marginally higher. The level of agreement of the CN lantern and the ColorDx D15 was identical for the number of crossings and C-index.

Because the CCT elliptical length and area are correlated, the agreement values using these parameters with the CN lantern were similar, and so only the CCT elliptical area was included in the figure.

The F-D15 and ColorDx D15 were the only tests at the 4.6m distance where the area under the ROC was significantly greater than chance, but the AC1 values of approximately 0.4 indicate that there was only a fair level of agreement between the D15s and the CN lantern at this distance. The agreement coefficients for these two tests were similar at the other test distances with a slight increase at the 0.57 m test distance. None of the arrangement tests agreement values were statistically different at the various viewing distances or from each other based on the 95% confidence interval. Note that the C-index and number of allowable crossings on the Farnsworth D15 increased as the viewing distance decreased, but the values were constant for the ColorDx D15.

At the closer test distances, the area under the ROC curves was greater than chance for the HRR, ColorDx PIP and CAD. The CCT area under the ROC was greater than chance only at the 0.57 m distance. At the 2.3 m distance, the agreement values for the HRR,

ColorDx, and CAD were at least 0.70, which indicates good agreement with the CN lantern. The values for the HRR and ColorDx PIP were significantly higher than the arrangement test values, whereas the CAD value was not statistically different from the HRR and ColorDx PIP. The agreement values decreased as the viewing distance decreased further. At the closest viewing distance, the AC1 values for the HRR and ColorDx PIP were significantly lower than their values at the 2.3 m viewing distance; however, their values at the 0.57 m were not significantly different from those of the arrangement tests.
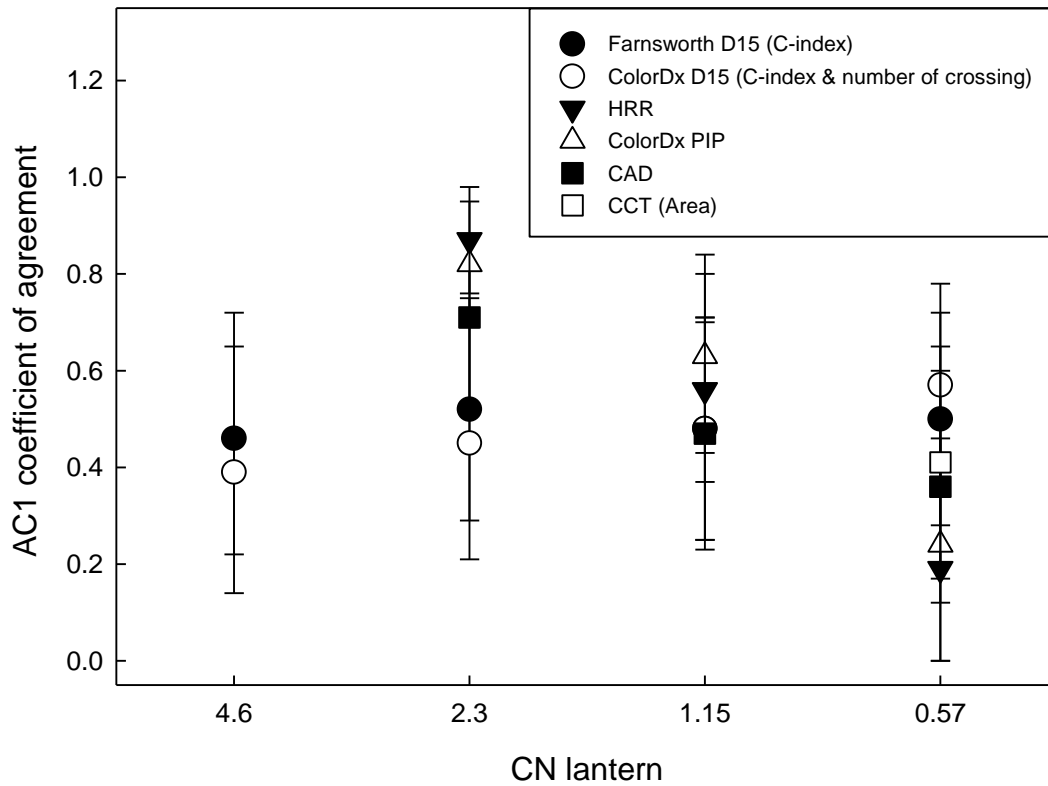
Figure 55. The AC1 coefficient of agreement values for the various tests with each distance at the CN lantern. The error bars are the 95% confidence intervals.

Table 28. The cut-off values for tests where the ROC curve area was significantly greater than 0.5 at different viewing distances.

	4.6	2.3	1.15	0.57
Farnsworth D15 (no. crossing)	3.167	3.167	3.833	3.833
Farnsworth D15 (C-index)	2.017	2.017	2.483	2.483
ColorDx D15 (C-index)	2.322	2.322	2.678	2.678
ColorDx D15 (no. crossing)	4	4	4	5.5
HRR (severity)		Mo or S*	Mo or S*	Mo or S*
ColorDx PIP (severity)		Mo or S*	Mo or S*	Mo or S*
CAD (Threshold)		14.47	22.45	23
CCT (area)				$1.12*10^{-3}$

*Mo is moderate and S is severe defect.

Figure 56 shows the PreF (top) and PreP (bottom) values for the tests at each distance. The general trend in the results was that the PreF values decreased with decreasing the viewing distances, whereas the PreP values showed the opposite trend. At the two furthest viewing distances, exceeding the cut-off score for the various tests was nearly perfect in predicting a failure on the CN lantern. Nevertheless, having a result better than the cut-off was generally poor in predicting a passing performance on the CN lantern. At the two shorter viewing distances (1.15m and 0.57), the PreP was equal 1.0 for the HRR and ColorDx PIP if the classification was moderate or severe. The trade-off was the reduction in the PreF values from nearly 1.0 at the longer distances to values ranging from 0.7 to 0.5. Thus, between 30%

and 50% of the individuals who were classified as moderate or severe by one of these clinical tests would pass the CN lantern at the two closest viewing distances.
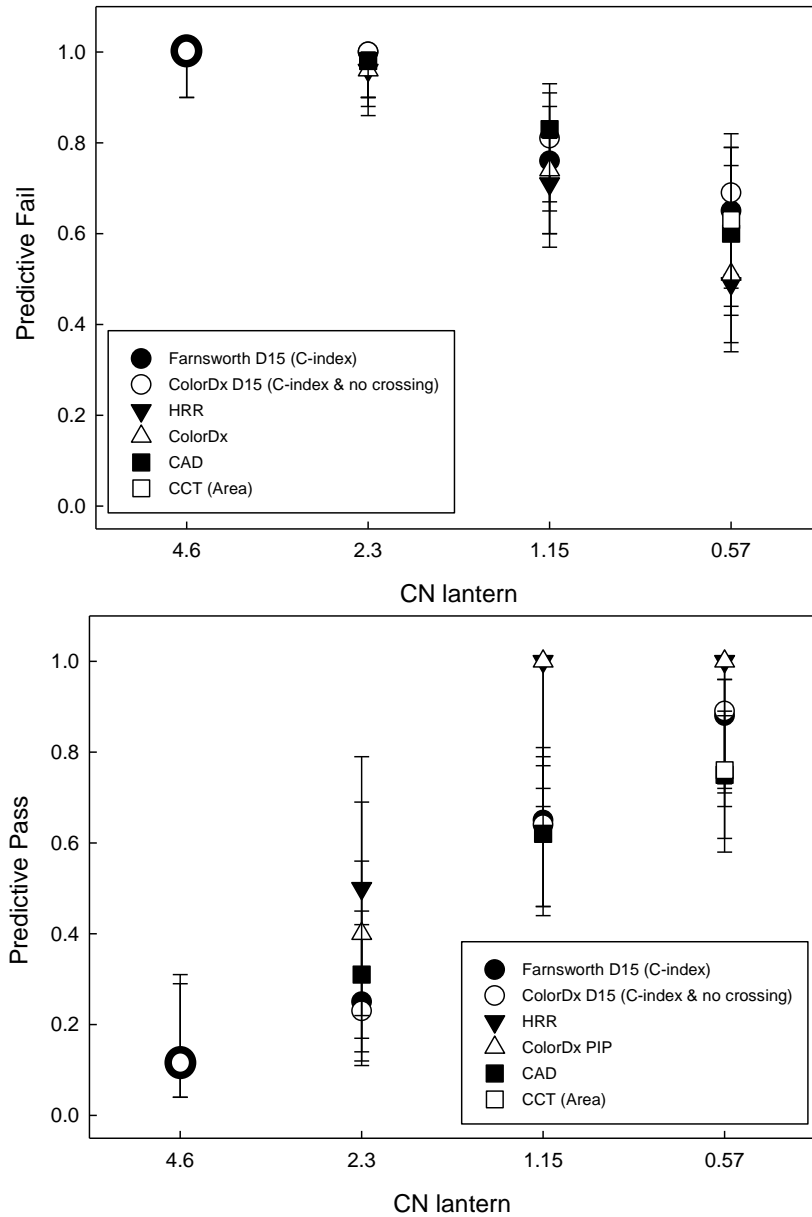
Figure 56. The predictive fail (top) predictive pass (bottom) values for the various tests with each distance at the CN lantern. The error bars are the 95% confidence intervals.

10.6 Discussion

The current colour vision standard for the major Canadian railway companies is that
individuals with a CVD must pass the CN lantern test if they wish to qualify for a position
requires them to operate equipment on the tracks. The results from this study may help in
counselling colour-defective patients who are interested in a career in the railroad industry.

The ColorDx D15 and Farnsworth D15 are excellent at predicting failures on the CN
lantern at the two longest distances. That is, exceeding the cut-off score of 4 crossings for
either version of the D15 essentially guarantees that the patient will fail the CN lantern at the
two longer viewing distances.

Predicting performance on the CN lantern at just the 4.6 m distance is difficult
because only 3 subject pass CN and fail other tests. Although they had mild CVD, their
results from the various clinical tests overlapped with the other subjects who failed at 4.6m.
Recall that railroads only administer the lantern if the candidate fails the Ishihara test and so
if the colour-defective passes the Ishihara, he or she will not have to take the lantern test. In
our study, three CVD subjects passed the Ishihara test, and so they would be classified as
normal and not subject to any additional testing. They also passed all other tests including the
lantern test at all distances so that in terms of passing colour vision tests found in most clinics
and identifying railway signal lights, they essentially have normal colour vision. Of the 60
subjects who failed the Ishihara test, only 3 passed the lantern test at 4.6 m. Thus, the
screening and threshold tests used in this study also have a PreP of 100% passing the CN
lantern at 4.6 m when their recommended pass/fail criteria is used. The PreF values for these

216

tests also approached 100% using the recommended pass/fail criteria. The reason why the area under the ROC curve did not reach significance for the threshold and qualitative tests was that none of these tests could identify the small number of CVD subjects who failed the Ishihara test, but passed the CN lantern at 4.6m.

Although being classified as having red-green CVD by a threshold or screening test has a high predictive value of failing the lantern at the 4.6 m, having more than 3 crossings on the F-D15 or more than 4 crossings on the ColorDx D15, virtually guarantees that the candidate will fail the CN lantern at 4.6 m and 2.3 m viewing distances. The difference in the cut-off point score between the F-D15 and ColorDx D15 is consistent with the previous finding in Chapter 7 that the ColorDx D15 is slightly easier to pass compared with the F-D15.

At the two closer viewing distances, the CVD pass rate on the CN lantern increased and a mild classification on the HRR and ColorDx PIP diagnostic plates was very effective in predicting who would pass with a PreP value of 100%. The decrease in the PreF values at this distance was due to a few dichromats who passed the lantern, but were classified as severe on the clinic tests. In terms of counselling CVD patients who might be interested in a restricted position, an assessment with either the HRR or ColorDx PIP diagnostic plates could be used, and a very mild or mild classification is a very good predictor of passing the CN lantern at the closer distances. A patient who has moderate-to-severe defect on these tests still has about 50% chance of passing the CN lantern at the closest distance; however, if they also fail one of the D15s, then the probability of passing decreases to 33%.

217

The somewhat surprising result was that the newer computer-based threshold tests and the RCCT were no better than chance in predicting the CN lantern results at the various distances. As mention previously, the poor predictive results at the 4.6 m distance were due to the fact that nearly all the CVD individuals who were classified as abnormal on these tests failed the lantern. At the other test distances, there was substantial overlap in the scores for individuals who passed and failed the CN lantern. Figure 57 shows an example. The graphs show the distribution of the ColorDx PIP qualitative severity (top) and LandC average of the L and M contrast thresholds (bottom) for subjects who passed or failed the CN lantern test at 1.15m. This distance and tests were selected because the AC1 agreement, PreP and PreF for the ColorDx PIP test were better than at the other distances and the area under the ROC curve for the LandC test was not significantly different from chance. Although some individuals with a severe classification on the ColorDx PIP did pass the CN lantern, none of the subjects who were classified as mild-to-moderate failed. This is why the PreP and AC1 values were relatively good. In contrast, the range of LandC thresholds for those who passed the lantern is nearly identical to the range for those who failed so that using any cut-off point on the LandC threshold will likely result in a chance agreement.
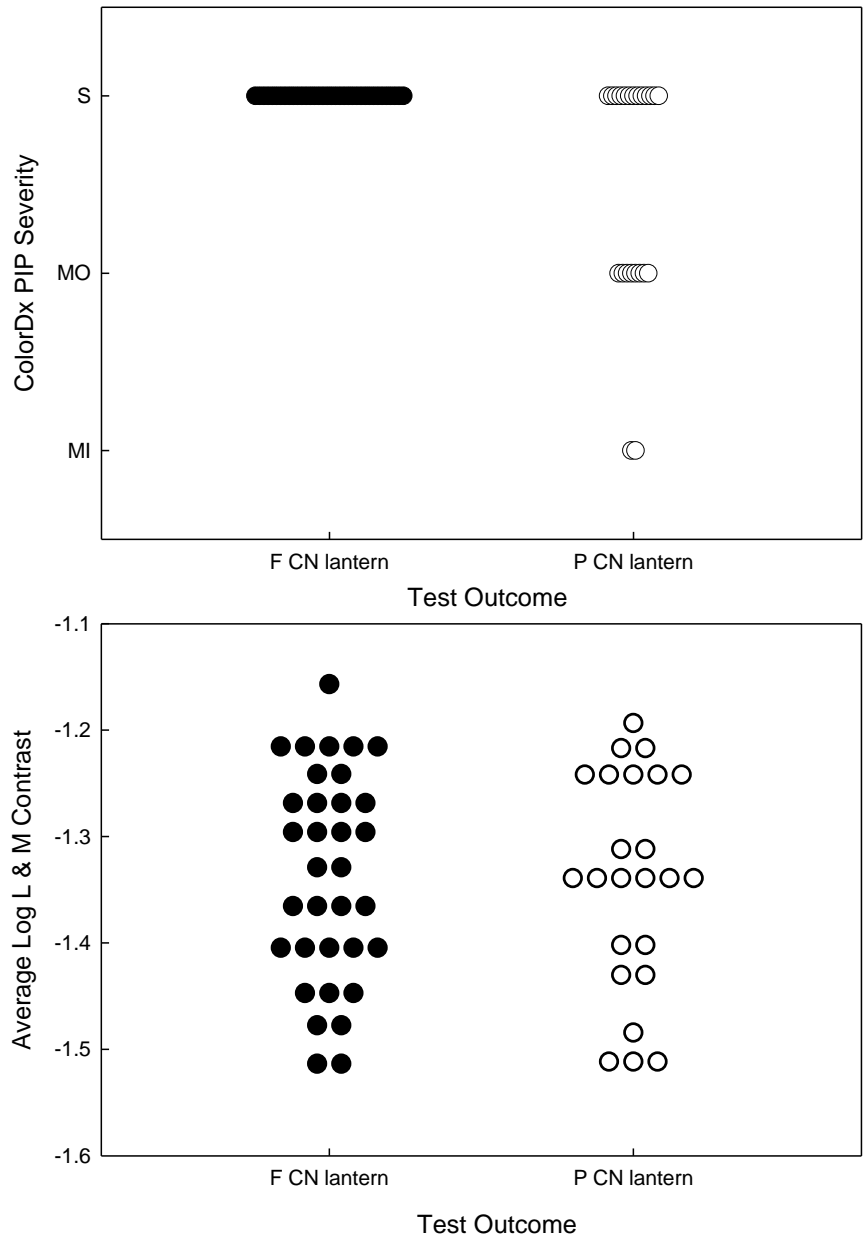
Figure 57. Dot histogram showing the distributions of the ColorDx PIP qualitative severity (top) and LandC average of the L and M cone thresholds (bottom) for subjects who passed or failed

the CN lantern at 1.15m. P-CN lantern is passed-CN lantern, and F-CN lantern is failed-CN lantern. MI is mild, MO is moderate and S is severe on the ColorDx PIP severity.

10.7 Conclusion

In order to counsel patients regarding a career as a locomotive engineer or conductor, a clinician should have either the HRR or ColorDx PIP along with a D15 test. If a patient has a colour vision defect and the defect was missed by one of these screening tests, then they have nearly a 100% probability of passing the CN lantern at all distances. It is likely that they would also pass the Ishihara test at the railway medical screening. Failure on the screening series is highly predictive of failing the CN lantern at the 4.6 m viewing distance, and if they also fail the D15, then a failure on the CN lantern at this distance is essentially guaranteed, and so they would not likely qualify for an unrestricted position. If they are applying for a position in the yard where the sighting distances are shorter, a mild-to-moderate classification on HRR or ColorDx PIP indicates that the patient has a high probability of passing at the two closer viewing distances. A severe classification does not rule out the possibility of pass the lantern at these distances, but a severe classification and more than 3 crossings on the D15 indicates that patient has only a 33% chance of qualifying for a position with shorter sighting distances.

Chapter 11

Colour Vision Defectives Experience with Colours

11.1 Introduction

Most individuals with a congenital red-green CVD report problem in making colour-related judgements in their daily lives (Steward and Cole, 1989; Tagarelli et al., 2004; Cole and Steward, 1997). These difficulties include having problems in selecting the appropriate colour of clothes, furniture, fruits, flowers and accessories or distinguishing the colour of materials, wires and paints (Steward and Cole, 1989; Tagarelli et al., 2004; Cole and Steward, 1997). The problems in daily life activities occurred more frequently in dichromats compared with anomalous trichromats (Steward and Cole, 1989). In contrast, only 2% of the colour-normals reported difficulties in making colour-related judgments.

Although there is general agreement that CVDs are more likely to experience problems with daily colour judgments, surveys differ in the frequencies who report problems. Steward and Cole (1989) reported that 73.5% of the CVDs have problems whereas Tagarelli et al. (2004) survey of CVNs and CVDs in Southern Italy found that only 50% of the CVDs reported problems with colour judgments in their everyday life tasks. The percent of CVDs who reported difficulties in selecting colour clothes, accessories, cars, paints, furniture, and wallpaper in their survey was 24%, which was 1/3 of the percentage reported by Steward and Cole (Tagarelli et al.; 2004). The frequency of colour-defectives in Tagarelli et al's suvery

who reported difficulty in distinguishing the colour of traffic light was 5%, which was also lower than the 29% reported by Steward and Cole. The lower frequencies reported by Tagaerlli et al. were likely due to differences in the range of ages. Tagaerlli et al survey university students who may not have encountered as many situations where colour judgments may have impacted their performance, whereas Steward and Cole surveyed individuals from 11 to 65 with a mean age of 40. Because the subjects in Steward and Cole's study were older, they may have been exposed to more situations were colour judgment occurred; for example driving.

Nevertheless, some studies reported that the CVDs had minimal problems in making colour judgments. Miles and Carige (1931) study reported that 27 colour-defective salesmen whose work involved coloured goods had few problems with colour or had to use secondary clues to help with their colour judgments (Miles and Carige, 1931). It is uncertain whether there was any self-selection within this sample or whether the colours were predominantly ones that were not confused by colour-defectives.

As part of this project comparing on the clinical utility of a variety of colour vision tests, we administered a short questionnaire to participants to determine whether colour-defectives drawn from a primarily university environment reported similar problems regarding colour-related tasks.

11.2 Methods

The demographics of the subjects are described in Chapter 3, and the questionnaire is in Appendix 3. All 60 CVNs 68 CVDs responded.

11.3 Results

None of the subjects reported any vision problems or eye diseases other than wearing corrected lenses, having a colour vision deficiencies or having a dry eye. All the subjects with normal colour vision were certain about their normality of colour vision. Nearly, all the CVD participants (99% of the CVD sample) were aware of their colour vision problem. The single exception was a 17 yrs old male who was unsure as to whether he had a colour vision defect or not. He was motivated to participate in the study to find out if he had colour vision deficiency. For those CVD participants who had prior knowledge of their defect, they were asked when they first became aware of their colour vision deficiency. Figure 58 shows the results using the US National Library of Medicine, National Institute of Health timeline (U.S. National Library of Medicine, 2018). Over 50% were aware of their defect by the age of 12 yrs, and over 90% were aware of their defect by the time they graduated from high school.
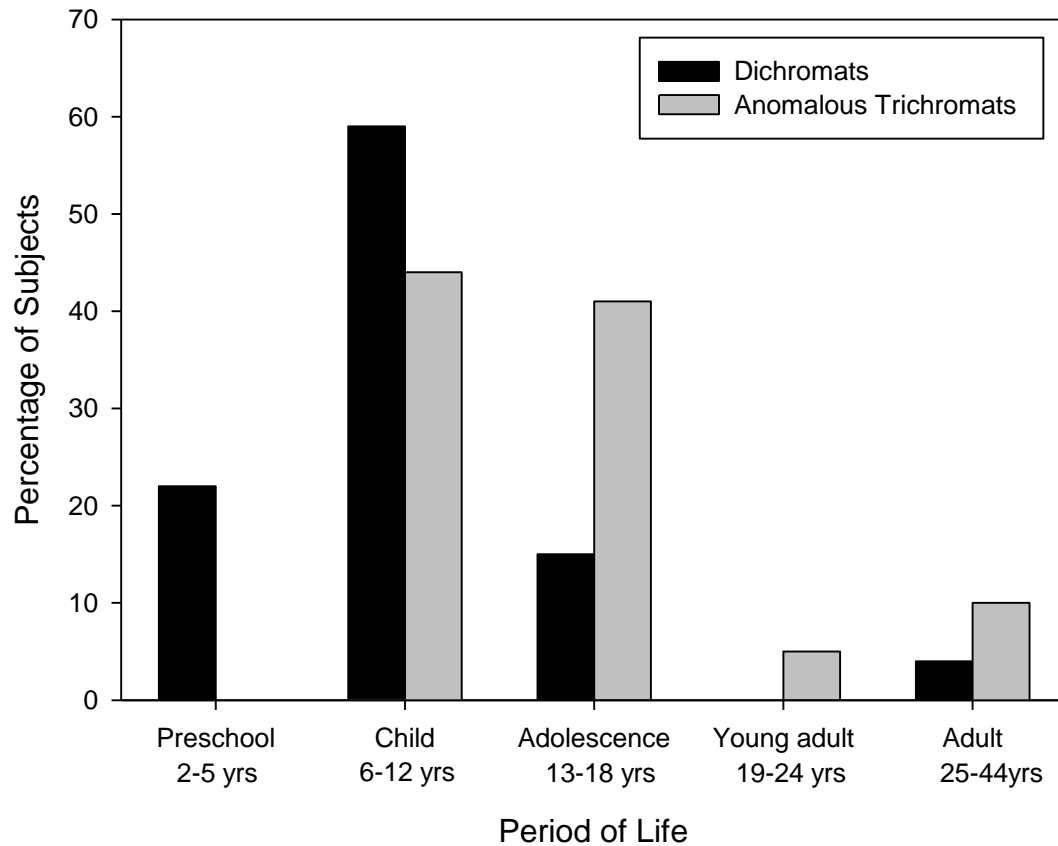
Figure 58. The percentage of CVDs (Dichromats and Anomalous Trichromats) who became aware of colour vision deficiency at each stage of life. Preschool (2 yrs to 5 yrs), Child (6yrs to 12 yrs), Adolescence (13yrs to 18 yrs), young adult (19yrs to 24 yrs), adult (25yrs to 44 yrs) (U.S. National Library of Medicine, 2018).

Figure 59 shows who informed the colour-defectives of their defect. An optometrist informed 47% of the CVDs with 4.5% (3 protanopes and one protanomalous) reporting that

they went to the optometrist specifically for colour vision test because family members or teachers noticed their problems with colour identification. Although they were sure about the defect, 12% did not report any independent confirmation. The remaining 4.5% (n=3) were informed by a nurse, aunt, or Civil Aviation Authorities.
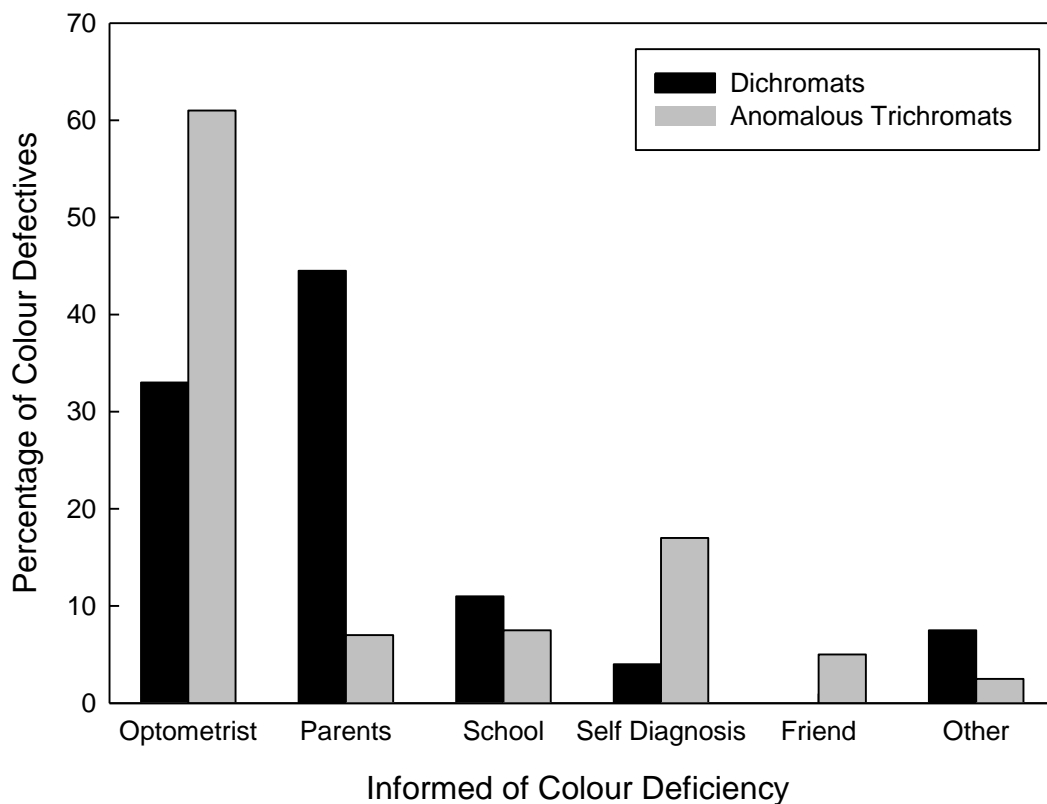
Figure 59. Percentages of how the CVDs became aware of their colour vision deficiency. Other, indicates that they were informed by a nurse, aunt and Civic Aviation Authorities.

11.3.1 Difficulties with colour related to their daily activities

The participants were asked some general questions about their experience with colour judgments. Table 29 shows the percentage of subjects who answered yes to each question. The percentage of dichromats who reported difficulties was more than double of the percentage of anomalous trichromats. Their responses were varied as to the major colour-related problem that they encountered. Forty-three percent of the CVDs reported having difficulty with distinguishing the colour of clothes, fruit, websites colour codes, signal lights, and colour balancing photos. Ten percent of the CVDs reported the colours that they confused without being prompted. These colours were distinguishing shades of red and blue with purple; red with green; red with yellow; brown with green and yellow with lime.

The participants were questioned regarding difficulties in detecting or recognizing the colour signals of the traffic light. Overall, 10% of sample reported problems identifying traffic signal light colours. This was more common in the dichromatic group with 18.5% of the dichromats reporting this problem compared with a 5% of the anomalous trichromats. More subjects reported problems locating the traffic signal lights at night when numerous other street lights are surrounding the traffic signal or in the background. The overall percentage was 25% with 33% of the dichromats reporting this problem and 20% of the anomalous trichromats.

Thirty percent of the dichromats and 2.5% of the anomalous trichromats reported that their color vision defect did influence their career choice. Overall, these individuals represented 13% of our sample.

Table 29. The percentage of subjects who responded to each question with yes response. Current, S&C and T et al. represent our study, Steward & Cole (1989) study, Tagarelli et al. (2004) study respectively. The numbers in parentheses are the number subjects in each category. The black rectangles indicate that the question was not asked and NA means that the colour-normals were not asked that question.

Question	Normals			Dichromats		Anomalous Trichromats		Protans		Deutans		All colour defectives		
	Current (60)	S&C (102)	T et al (302)	Current (27)	S&C (37)	Current (41)	S&C (65)	Current (28)	S&C (36)	Current (40)	S&C (66)	Current (68)	S&C (102)	T et al (151)
Do you have any difficulty with colour judgments in your daily activities? If yes, what is the major problem	0	4	▆	66.5	86	27	66	57	0	32.5	11	43	73.5	▆
Do you have any difficulty identifying the colour of traffic lights?	0	0	2	18.5	49	5	18	14	39	7.5	24	10	29	4.8
Do you have any problem finding traffic signal lights at night when there are numerous other street lights surrounding the traffic signal or in the background?	0	2	▆	33	33	19.5	31	28.5	36	22.5	21	25	31	▆
Did your colour vision problem affect your career choice?	NA	NA	NA	30	43	2.5	29	25	38	5	33	13	34	0

The main purpose of the F-D15 is to separate individuals who are likely to encounter problems in making colour judgments in their daily lives from those who are not likely to encounter problems. Although all but one the dichromats failed the F-D15, approximately 34% did not report any problems with colour judgments in their daily lives. This percentage was slightly higher for the anomalous trichromats with 46% of those who failed the F-D15 reporting no difficulties with colour-related tasks. However, 17% of the anomalous trichromats who passed the F-D15 did report problems in making colour judgments in their daily lives.

Previous research showed that identifying the colour of objects by individuals with CVD can be idiosyncratic and may depend on the context (Hovis et al., 1994; Ramaswamy & Hovis, 2011). One way to show how context can influence their colour naming was to ask a question about the colour of the man figure in the pedestrian traffic light indicating that it is safe to cross the street. In North America, the figure is white and likely to appear similar in colour to the green traffic signal for colour-defectives because the green traffic lights are near the line of confusion with white (CIE,1980).

All of the CVNs identified the figure correctly as white, whereas 53% of the CVDs identified the colour as green and 3% reported that the figure was yellow. Figure 60 shows the percentages of the different colour vision defectives who identified the figure as green. This mistake was slightly less likely to occur for the deutans and anomalous trichromats. Interestingly, 62% of the CVDs (12 anomalous trichromats and 6 dichromats) who reported no difficulties in making colour judgments reported that the colour of the figure was green. In addition, one of the two subjects who reported the colour was yellow also reported that he had no problems with colours.
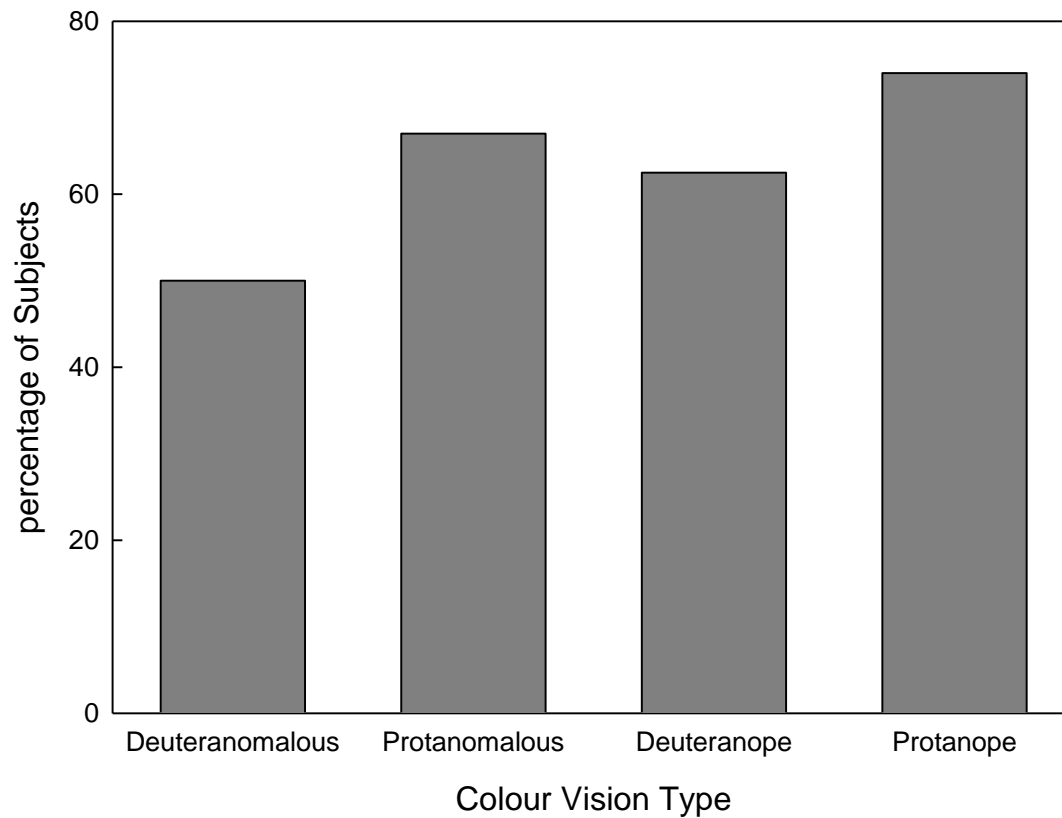
Figure 60. The percentage of each type CVDs on each type of colour-defective vision deficiency who reported the man figure on the pedestrian traffic light as green.

11.4 Discussion

The result that essentially all of the CVDs were aware of their colour vision problem was expected because the recruitment specifically targeted individuals with a colour vision defect. Steward and Cole (1989) surveyed patients in private practices and Tagaerlli et al. (2004) survey university students.

Of the participants who were certain about their defect, 92.5% of the dichromats and 85% of the anomalous trichromats were aware of their colour vision deficiency by the time they graduated from high school. This was larger than 70% of the dichromats and 36% of the

anomalous trichromats reported by Steward and Cole (1989) and the difference is likely due to our targeted recruitment.

The percentage of the CVDs who had trouble with colour judgments on a daily basis was 43%. This was smaller than the 73.5% of the CVDs who reported more than one difficulty in the Steward and Cole's (1989) study. The reason for the large difference may be a result that we asked a very general question and may not have triggered their memory, as would questions about a specific colour-related tasks asked by Steward and Cole (1989). Of course, our subjects may have also learned to avoid difficult colour judgments.

Our study found that only 18.5% of the dichromats and 5% of the anomalous trichromats recalled difficulty in identifying the colour of traffic lights, which was smaller than 49% of the dichromats and 18% of the anomalous trichromats found by Steward and Cole (Steward & Cole, 1989). The difference was likely a result of their subjects being older and not primarily university students and so their subjects may have had more cumulative driving experience than our subjects. The results from Tagarelli et al. (2004) study that only 5% of the CVDs, who were a university students level, reported difficulties in identifying traffic signal lights supports the hypothesis that cumulative driving experience is responsible for the difference between the results of this study and Steward and Cole's findings. However, it could have also been how the question was phrased in the Tagarelli et al. study. They asked whether the participants believed that they could stop at a red signal light if the order of the colours of the traffic light were changed.

Although the percentage of CVDs who reported problems identifying traffic lights was lower than reported by Steward and Cole, the ratio of the deutans to protans in this subgroup was very similar to their findings. Our deutan to protan ratio was 1:1.86 compared

with 1:1.63 reported by Steward and Cole (1989). This confirms their finding that protans have more difficulty in recognizing signal lights (Cole, 2002).

Similar to Steward and Cole, the CVDs were more likely to report that they had difficulty distinguishing the traffic signal lights from the background street lighting rather than identifying the colour of the traffic light itself. Thirty-three percent of dichromats reported confusing traffic light with streetlights in both studies, whereas only around 20% of anomalous trichromats reported that difficulty in our study compared with 31% reported by Steward and Cole (1989). Again, the difference in the percentage of anomalous trichromats could be due to driving exposure.

The lower percentage of CVDs reporting that their defect influenced their career choice in our study may also be because we were drawing from mainly a university-based community and factors, other than their colour vision defect, determined their occupation or major program. It is also possible that the subjects were unaware that they self-selected their particular occupation or major based on their colour vision defect. The result that there was also a low percentage of subjects who reported problems with colour judgments suggests that they may have chosen daily activities that did not require frequent colour-related judgments. Interestingly one mild deuteranomalous individual reported that his colour vision defect did influence his career path. This individual reported the problem in finding the right colour balance for his photography. His inability to obtain the correct colour balance kept him from pursuing a career in video or photography. What makes this subject so

interesting is that he actually passed all of the screening tests, but had a Rayleigh match that was clearly outside the normal values with a matching range of 2, and so it is likely that he would have never known why he had difficulty with colour balance if he had not participated in this study.

The result that a slight majority of colour vision defectives who believed that the man-figure on the pedestrian traffic light was green is a demonstration of how context can influence their colour naming. Because they were asked to identify the colour based on memory, we do not know if the same results would occur if they were viewing the actual lights. We suspect that the percentage would actually be higher because in traffic light context the green colour indicates it is safe to cross the street. Two CVD subjects who reported that the figure was white mentioned that they recently learned that the light was white, but they still believed it is green. The relative high percentage of CVD who reported that the figure was green and also reported that they did not have any difficulties in making colour judgments suggests that they could not remember any problems or that no one informed them of their colour identification errors because their behavior to the signal light was correct.

Most color memory research used CVN subjects. These studies have shown that the colours are easier to discriminate when they presented simultaneously, whereas the variability is higher when the discrimination is done from memory (Zhu et al., 2017; Bloj et al., 2016; Pérez-Carpinell et al., 1998; and Newhall et al. 1957). The only study using CVD subjects was the work by Pérez–Carpinell et al. (2001). They asked anomalous trichromats subjects to match colours simultaneously and successively (based on memory) for a selected

colours from a paint company's glossy colour chart. Their results showed that the mean

colour differences obtained from the simultaneous colour match were lower than if the match

based on colour memory. This indicates that colour can be discrimnated easily when it

presented simultaneously compared to the discrimination based on colour memory. Further

study to expand on how context influences colour memory and colour judgments in CVDs

would be interesting.

11.5 Conclusion

The study showed that some individuals with CVD do report some difficulty in

distinguishing colour in their daylily life activities. As expected, dichromats are more likely

to have difficulty in identifying colours in their daily lives. Context can play a key role in

how CVDs label colours.

Chapter 12

Conclusion

Transforming the threshold data of the colour vision tests for CVNs and CVDs into a common colour space simplified the comparison between tests. The CCT ellipse area results were smaller than the CAD results. This was likely because the CAD measure thresholds in the region of color space where subjects would expect to have a larger thresholds along with fatigue or attention that the subjects might have as CAD took longer to complete than CCT. Because the CAD samples near the dichromatic lines of confusion, the CVDs thresholds were higher in these directions and so the areas of ellipses were larger than the CCT, which sampled fewer locations that were more likely to be slightly away from a line of confusion. Because of the sampling strategy, the angle of the CAD ellipse had a higher validity as to the type of defect. These results might not hold if all 3 of the possible ellipses for the CCT were used.

The RCCT, LandC, and CCT Tri tests measured, or estimated, thresholds in only 3 directions on, or near one of each of the dichromatic lines of confusion. The results for the CVN were that the L and M cone thresholds were the largest on the CCT Tri followed by the RCCT and smallest on LandC. The differences between tests were significant. For the CVDs, the RCCT and LandC tests had higher cone threshold on the corresponding type of defect and near normal threshold on the other cone. That is protans had a high threshold on the L cone and near normal threshold on the M-cone. Similarly, the deutans had higher M cone

thresholds and near normal thresholds on the L cone. The CCT Tri was the exception as it misclassified a large number of protans as deutans and vice versa.

The F-D15 and ColorDx D15 colorimetric properties were evaluated in terms of how well the predicted dichromatic arrangements matched the actual arrangements. The colormetric model for the F-D15 and ColorDx D15 tests showed good agreement with dichromatic and CVNs arrangements on both tests. However, both the predictions and results showed that deuteranope subjects had fewer crossings on the ColorDx D15 test, which suggests that the ColorDx D15 would be easier to pass for deutans. This could be a result of the larger sized coloured stimuli in the ColorDx D15. However, protanope subjects had more crossings than the predicted on both F-D15 and ColorDx D15.

The repeatability of the F-D15 and ColorDx D15 for either the 1st trial or 2/3 trials using two different pass/fail criteria showed that there were minimal practice effects within a session. This result indicates that there was not a strong advantage for requiring a patient to pass on 2/3 trails versus using the 1st trial only for either the ColorDx D15 or F-D15; however, this was conditional on not having any additional stress, such as trying to qualify for a job. The only advantage of asking the patients to perform the test multiple times within a session was that it showed slightly better between-session repeatability for the classification, particularly for the F-D15. However, this improvement was more likely due to improvement of the examiner in interpreting the results rather than change in the subject's arrangement. Using a failure criterion of >1 crossing is recommended because it had slightly better between-session repeatability for both pass/fail and classification of the type of defect

235

compared with the >0 crossing criterion. The ColorDx D15, which is the computerized version of F-D15, has good agreement with F-D15 and could be an adequate substitute to the F-D15.

One of the aim of this thesis was to determine which one of the next generation of colour vision tests could be used as a substitute for some current occupational colour vision tests. The occupational tests were the F-D15, HWA, and CAD. The results showed that the ColorDx D15 test would be the best option to replace the F-D15. The CAD would be the next best option, but 31% of the individuals had discrepancies between the two tests. The newer computer tests had a reasonable level of agreement and high predictive values for failure because most of the CVDs (92%) failed the HWA. In comparison with CAD-pilot criterion test, none of the other tests could adequately predict performance on the CAD-pilot colour vision test. The low agreement with other tests was likely due to the small number of CVD subjects who passed the CAD test in our study.

Because sighting distances for railway signal lights can vary, individuals with a colour vision defect who do not qualify for positions requiring longer sighting distances may qualify for positions using shorter sighting distances. The error rate on the CN Lantern test decreased with decreased viewing distance. The major change is when the viewing distance decreases from the maximum of 4.6 m to 2.3 m. The increase in retinal illumination at the shorter distances is likely responsible for the improvement in performance, but it is likely that the improved performance asymptotes when the retinal illuminance is 2.3 log units above the CVN threshold. Performance on the CN lantern improves with practice; however,

236

the reduction in the number of errors affected the pass/fail repeatability only at the closest viewing distance. The study confirmed that 65% of the CVDs could carry the work on the railyard safely with some restriction at shorter viewing distance when the sighting distance is within 82 m.

Because the CN Lantern test is limited in its availability, a variety of different clinical colour vision tests were investigated to determine whether they would be useful in predicting performance on the lantern. This information would help clinicians in their counselling of their CVD patients who may be interested in a career in the railroad industry. The results indicate that the HRR or ColorDx PIP along with a D15 test would be useful for this purpose. Individuals with colour vision deficiency who pass the screening series would have 100% probability of passing the CN Lantern at all distances and failing the screening series is highly predictive of failing the CN lantern at a viewing distance of 4.6m. If the person also fails one of the D15s, then it is nearly certain that they will fail CN lantern at this distance. If the position was working in the yard where the sighting distances are shorter and so is the CN Lantern viewing distance, then a classification of mild-to-moderate on HRR or ColorDx PIP indicates that the patient has a high probability of passing at the two closer viewing distances. If the candidate was classified with a severe defect and more than 3 crossings on the F-D15, he/she still has a chance of 30% of qualifying for a position at shorter sighting distances.

Colour-defective individuals reported problems regarding colour related judgments in their daily activities that were similar to problems reported in other studies, but the frequency was lower. As expected, dichromats reported difficulties more frequently than the anomalous trichromats. A slight majority of CVDs believed that the man-figure on the pedestrian traffic light was green in colour. This is a demonstration of how context can influence their colour naming.

To conclude, the next generation colour vision tests have some advantages that would benefit the users. The LandC and CAD tests can measure the chromatic threshold for CVN and CVDs, but these two tests were not perfect at predicting performance of the CN lantern test. However, the CAD and LandC tests had a very good level of agreements with the HWA. The ColorDx D15 and CAD tests were sufficient substitutes of the F-D15 with the ColorDx D15 being slightly easier to pass. The discrepancies between the CAD and F-D15 were more equally distributed between subjects that fail one and pass the other. The RCCT test was a good test to distinguish the CVNs from the CVDs; however, it cannot measure the threshold for CVNs. The RCCT takes the shorter time to complete the task in comparison with the other computer-based tests, and it can screen for both blue-yellow and red-green defects. The computerized tests do not require special lighting. More importantly, the tests were designed to be difficult to memorized and have less administrator bias. However, the monitors have to be calibrated on a regular basis in order to ensure that the proper colours are displayed. The CCT discrimination ellipse and CCT Tri tests may not be a suitable test if clinician needs to

classify the type of red-green defect as both tests misclassified a number of protans as

deutans and vice versa.

Bibliography

Almustanyir, A., & Hovis, J. K. (2015). Military Research ColorDx and Printed Color Vision Tests. *Aerospace Medicine and Human Performance, 86*(10), 852-859

Atchison, D. A., Bowman, K. J., & Vingrys, A. J. (1991). Quantitative scoring methods for D15 panel tests in the diagnosis of congenital color vision deficiencies. Optom Vis Sci, 68(1), 41-48.

Barbur, J. L., Harlow, A. J., & Plant, G. T. (1994). Insights into the Different Exploits of Colour in the Visual Cortex. Proc.R.Soc.Lond.B, 258(1353), 327-334.

Barbur, J. L. (2004). 'Double-blindsight'revealed through the processing of color and luminance contrast defined motion signals. *Progress in brain research*, *144*, 243-259.

Barbur, J. L., Rodriguez-Carmona, M., & Harlow, J. A. (2006). *Establishing the statistical limits of "normal" chromatic sensitivity.* Paper presented at the CIE Proceedings Expert Symposium – 75 Years of the CIE Standard Observer Ottawa. May 2006:06-75.

Barbur, J. L., & Rodriguez-Carmona, M. (2015). Color vision changes in normal aging.

Belcher, S. J., K. W. Greenshields, and W. D. Wright (1958):. Colour vision survey: using the Ishihara, Dvorine, Boström and Kugelberg, Boström, and American-Optical Hardy-Rand-Rittler tests The British journal of ophthalmology 42.6 ,355.

Birch, J., Barbur, J. L., & Harlow, A. J. (1992). New method based on random luminance masking for measuring isochromatic zones using high resolution colour displays. *Ophthalmic and Physiological Optics*, *12*(2), 133-136.

Birch, J., & McKeever, L. M. (1993). Survey of the accuracy of new pseudoisochromatic plates. Ophthalmic and Physiological Optics, 13(1), 35-40.

Birch J. (1997). Efficiency of the Ishihara test for identifying red-green colour deficiency. Ophthalmic Physiol Opt.17(5):403-408.

Birch, J. (2008a). Pass rates for the Farnsworth D15 colour vision test. Ophthalmic and Physiological Optics, 28(3), 259-264.

Birch, J. (2008b). Performance of colour-deficient people on the Holmes-Wright lantern (type A): consistency of occupational colour vision standards in aviation. Ophthalmic Physiol Opt, 28(3), 253-258.

Birch, J. (2010). Identification of red–green colour deficiency: sensitivity of the Ishihara and American Optical Company (Hard, Rand and Rittler) pseudo-isochromatic plates to identify slight anomalous trichromatism. *Ophthalmic and Physiological Optics, 30*(5), 667-671.

Bloj, M., Weiß, D., & Gegenfurtner, K. R. (2016). Bias effects of short-and long-term color memory for unique objects. *JOSA A*, *33*(4), 492-500.

Brainard, D. H. (2003). Color appearance and color difference specification. *The science of color*, *2*, 191-216.

Brettel, H., Viénot, F., & Mollon, J. D. (1997). Computerized simulation of color appearance for dichromats. *JOSA A*, *14*(10), 2647-2655.

Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of clinical epidemiology*, *43*(6), 551-558

CIE. International recommendations for color vision requirements for transport. Report 143. International Commission on Illumination, CIE technical report, Vienna, Austria. 2001.

Civil Aviation Authority, Safety Regulation Group. Minimum color vision requirements for professional flight crew - Part 2: Task analysis. Paper 2006/04. 2006b.

Civil Aviation Authority, Safety Regulation Group. Minimum Colour Vision Requirements for Professional Flight Crew Recommendations for new colour vision standards. Paper 2009/04. 2009

Cole, B. L., & Brown, B. (1966). Optimum intensity of red road-traffic signal lights for normal and protanopic observers. *JOSA*, *56*(4), 516-522.

Cole, B. L., & Vingrys, A. J. (1982). A Survey and Evaluation of Lantern Tests of Color Vision. Am J Optom Arch Am Acad of Optom, 59(4), 346.

Cole, B. L., & Vingrys, A. J. (1983). Who fails lantern tests? Doc Ophthalmol., 55(3), 157-175.

Cole, B. L. (1963). Misuse of the Ishihara test for colour blindness. *The British journal of physiological optics*, *20*, 113-118.

Cole, B. L., & Steward, J. M. (1997). Some (but only a few) colour vision defectives have no difficulty with colour. *John Dalton's Colour Vision Legacy. London: Taylor and Francis*, 235-239.

Cole, B. L. (2002). Protan colour vision deficiency and road accidents. *Clinical and experimental optometry*, *85*(4), 246-253.

Cole, B. L., Lian, K. Y., & Lakkis, C. (2006a). The new Richmond HRR pseudoisochromatic test for colour vision is better than the Ishihara test. *Clin Exp Optom., 89*, 73-80.

Cole, B. L., Lian, K. Y., Sharpe, K., & Lakkis, C. (2006). Categorical color naming of surface color codes by people with abnormal color vision. *Optometry & Vision Science*, *83*(12), 879-886.

Commission Internationale de l'Eclairage. (1980). Light signals for road traffic control (CIE Publication 48). Vienna: Author.

Dain, S. J., & Adams, A. J. (1990). Comparison of the standard and Adams desaturated D-15 tests with congenital colour vision deficiencies. *Ophthalmic and Physiological Optics*, *10*(1), 40-45.

Dain, S. J., Casolin, A., & Long, J. (2015). Color vision and the railways: Part 2. Comparison of the CN Lantern used on the Canadian Railways and Railway LED lantern tests. *Optometry & Vision Science*, *92*(2), 147-151.

Deeb, SS. (2004). Molecular genetics of colour vision deficiencies. *Clin Exp Optom., 87*(4-5), 224-229.

Department of Transportation, Federal Railroad Administration (2015) Best Practices for Designing Vision Field Tests for Locomotive Engineers or Conductors. 49 CFR Parts 240 and 242, Washington, DC.

Farnsworth, D. (1947). The Farnsworth Dichotomous Test for Color Blindness Panel D-15. New York: The Psychological Corp.

Fletcher, Robert, and Janet Voke (1985). *Defective colour vision, fundamentals, diagnosis and management*. CRC Press.

243

Fiorentini, A., Porciatti, V., Morrone, M. C., & Burr, D. C. (1996). Visual ageing: unspecific decline of the responses to luminance and colour. *Vision research*, *36*(21), 3557-3566.

Fu, C., Xiao, K., Karatzas, D., & Wuerger, S. M. (2009). Changes in colour perception with ageing. *Paper submitted to the AIC*.

Golz, J., & MacLeod, D. I. (2003). Colorimetry for CRT displays. *JOSA A*, *20*(5), 769-781

Gunkel RD, Cogan DG. (1978). Colorimetry by a new principle. *Arch Ophthalmol*; 96:331-334.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol, 61*.

Haskett, M.K., Hovis J.K. (1987)Comparison of the Standard Pseudoisochromatic Plates to the Ishihara color vision test. Optom Vis Sci 64: 211-216

Holmes, J. G. and Wright, W. D. (1982), A new colour-perception lantern. Color Res. Appl., 7: 82–88.

Hovis, JK, Almustanyir, A. Assessment of the Next Generation of Colour Vision Tests for Pilots and Aircrew. Submitted September 1, 2017 to M. G. Glaholt Defence Research and Development Canada,Toronto, ON 84 pages

Hovis, J. K., Cawker, C. L., & Cranton, D. (1990). Normative data for the Standard Pseudoisochromatic Plates--Part 2. *J Am Optom Assoc, 61*(12), 913-920.

Hovis, J. K., Cawker, C. L., & Cranton, D. (1996). Comparison of the Standard Pseudoisochromatic Plates--Parts 1 and 2--As screening tests for congenital red-green color vision deficiencies. *J Am Optom Assoc, 67*(6), 320-326.

Hovis, J.K, Ramaswamy, S., & Anderson, M. (2004). Repeatability indices for the Farnsworth D-15 test. Visual Neuroscience, 21(03), 449-453.

Hovis, J. K., Lu, B., & Neumann, P. (1994). The ability of colour vision tests to predict performance in the identification of wire colour. *Can J Optom*, *55*, 219-27.

Hovis, J.K, Ramaswamy, S., & Anderson, M. (2006). The effect of test distance on the CN lantern results." *Visual neuroscience* 23.3-4 (2006): 675-679

Hovis, J. K., & Oliphant, D. (1998). Validity of the Holmes–Wright lantern as a color vision test for the rail industry. *Vision research*, *38*(21), 3487-3491.

Hovis, J. K., & Oliphant, D. (2000). A lantern colour vision test for the rail industry. *American journal of industrial medicine*, *38*(6), 681-696.

Hovis, J. K. (2008). Repeatability of the Holmes-Wright type A lantern color vision test. Aviat Space Environ Med, 79(11), 1028-1033.

Hurvich, Leo M., and Dorothea Jameson. "Some quantitative aspects of an opponent-colors theory. II. Brightness, saturation, and hue in normal and dichromatic vision." *JOSA* 45.8 (1955): 602-616.

Jameson, Dorothea, and Leo M. Hurvich. "Some quantitative aspects of an opponent-colors theory. I. Chromatic responses and spectral saturation." *JOSA* 45.7 (1955): 546-552.

Jennings, B. J., & Barbur, J. L. (2010). Colour detection thresholds as a function of chromatic adaptation and light level. *Ophthalmic and Physiological Optics*, *30*(5), 560-567.

Kinney, J. A. S., Paulson, H. M., & Beare, A. N. (1979). The ability of color defectives to judge signal lights at sea. *JOSA*, *69*(1), 106-113.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. Vision Research, 39(16), 2729-2737

Lakowski, R. (1966). A critical evaluation of colour vision tests. *The British Journal of Physiological Optics, 23*(3), 186-209

Linhares, J. M., João, C. A., Silva, E. D., de Almeida, V. M., Santos, J. L., Álvaro, L., & Nascimento, S. M. (2016). Assessing the effects of dynamic luminance contrast noise masking on a color discrimination task. *JOSA A*, *33*(3), A178-A183.

Mollon, J.D. and Bowmaker, J.K. (1992). The spatial arrangement of cones in the primate fovea. Nature, 360, 677-9

Moreland, J. D. and Kerr, J. (1979) Optimization of a Rayleigh-type equation for the detection of tritanomaly. Vision Res. 19, 1369–1375

National Research Council (US) Committee on Vision. (1981). *Procedures for Testing Color Vision: Report of Working Group 41.* Washington (DC) National Academies Press (US): Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK217825/

Nathan, J., Henry, G. H., & Cole, B. L. (1964). Recognition of colored road traffic light signals by normal and color-vision-defective observers. *JOSA*, *54*(8), 1041-1045.

Oyster, Clyde W (1999). *The human eye: structure and function*. Sinauer Associates.

Paramei, G. V., & Oakley, B. (2014). Variation of color discrimination across the life span. *Journal of the Optical Society of America A, 31*(4), A375-A384.

Paramei, G. V. (2012). Color discrimination across four life decades assessed by the Cambridge Colour Test. *JOSA A*, *29*(2), A290-A297.

Paramei, G. V., Bimler, D. L., & Cavonius, C. R. (1998). Effect of luminance on color perception of protanopes. *Vision Research*, *38*(21), 3397-3401.

Pérez-Carpinell, J., Baldoví, R., de Fez, M. D., & Castro, J. (1998). Color memory matching: Time effect and other factors. *Color Research & Application*, *23*(4), 234-247.

Pérez–Carpinell, J., Camps, V. J., de Fez, M. D., & Castro, J. (2001). Color memory matching in normal and red-green anomalous trichromat subjects. *Color Research & Application*, *26*(2), 158-170.

Newhall, S. M., Burnham, R. W., & Clark, J. R. (1957). Comparison of successive with simultaneous color matching. *JOSA*, *47*(1), 43-56.

Pitt FHG. The nature of normal trichromatic and dichromatic vision. Proc R Soc Lond (Biol) 1944;132:101-17.

Pokorny,J. Smith, VC. Verriest, G., Pinckers. AJLG. (Eds.), Congenital and Acquired Color Vision Defects, Grune and Stratton, New York (1979)

Rabin, J. (2004). Quantification of color vision with cone contrast sensitivity. Vis Neurosci, 21(3), 483-485.

Rabin, J., Gooch, J., & Ivan, D. (2011). Rapid Quantification of Color Vision: The Cone Contrast Test. Investigative Ophthalmology & Visual Science, 52(2), 816-820.

Railway Association of Canada. (2016) Canadian Railway Medical Rules Handbook", Ottawa, Canada, https://www.railcan.ca.

Ramaswamy, S., & Hovis, J. K. (2011). Does dichromat color simulation predict color identification error rates?. *Optometry and Vision Science*, *88*(5), 621-627.

Regan B, Reffin J, Mollon J. (1994) Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. Vision Res;34:1279-1299

Remington, L. A. (2011). *Clinical anatomy of the visual system*. Elsevier Health Sciences

Rodriguez-Carmona, M., O'Neill-Biba, M., & Barbur, J. L. (2012). Assessing the Severity of Color Vision Loss with Implications for Aviation and Other Occupational Environments. *Aviation, Space, and Environmental Medicine, 83*(1), 19-29.

Rowland, L. S. (1943). Selection and Validation of Tests for Color Vision: Relationship between Degree of Color Deficiency and Ability to Identify Signals from a "Biscuit Gun.". *School of Aviation Medicine Project Report*, (137-7).

Rüfer, F., Sauter, B., Klettner, A., Göbel, K., Flammer, J., & Erb, C. (2012). Age-corrected reference values for the Heidelberg multi-color anomaloscope. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *250*(9), 1267-1273.

Shin YJ, Park KH, Hwang J, Wee WR, Lee JH. A new colour vision test to differentiate congenital and acquired colour vision defects. *Ophthalmology*. 2007;114(7):1341-1347.

Shinomori, K., Panorgias, A., & Werner, J. S. (2016). Discrimination thresholds of normal and anomalous trichromats: Model of senescent changes in ocular media density on the Cambridge Colour Test. Journal Optical Society of America A, 33(3), A65-A76.

Schneck, Marilyn E., et al. "Comparison of panel D-15 tests in a large older population." *Optometry and vision science: official publication of the American Academy of Optometry* 91.3 (2014): 284.

Schneck, M. E., Haegerstrom-Portnoy, G., Lott, L. A., & Brabyn, J. A. (2014). Comparison of Panel D-15 Tests in a large older population. Optom Vis Sci, 91(3), 284-290.

Schwartz, Steven (2009). *Visual Perception: A Clinical Orientation: A Clinical Orientation*. McGraw Hill Professional.

Seshadri, J., Christensen, J., Lakshminarayanan, V., & Bassi, C. J. (2005). Evaluation of the new web-based "Colour Assessment and Diagnosis" test. Optom Vis Sci., 82(10), 882-885

Sloan, L. L. (1944). A quantitative test for measuring degree of red-green color deficiency. *American Journal of Ophthalmology*, *27*(9), 941-949

Squire, T. J., Rodriguez-Carmona, M., Evans, A. D., & Barbur, J. L. (2005). Color vision tests for aviation: comparison of the anomaloscope and three lantern types. *Aviation, space, and environmental medicine*, *76*(5), 421-429.

Steward, J. M., & Cole, B. L. (1989). The effect of object size on the performance of colour ordering and discrimination tasks. In *Colour Vision Deficiencies IX* (pp. 79-88). Springer Netherlands.

Steward, J. M., & Cole, B. L. (1989). What do color vision defectives say about everyday tasks?. *Optometry & Vision Science*, *66*(5), 288-295.

Tagarelli, A., Piro, A., Tagarelli, G., Lantieri, P. B., Risso, D., & Olivieri, R. L. (2004). Colour blindness in everyday life and car driving. *Acta Ophthalmologica*, *82*(4), 436-442

U.S. National Library of Medicine. (2018). Medical Subject Headings 2018, Retrieved from www.ncbi.nlm.nih.gov/mesh/68000328

Vingrys, A. J., & Cole, B. L. (1983). Validation of the Holmes-Wright lanterns for testing colour vision. Ophthalmic and Physiological Optics, 3(2), 137-152.

Vingrys, A. J., & King-Smith, P. E. (1988). A quantitative scoring technique for panel tests of color vision. Investigative Ophthalmology & Visual Science, 29(1), 50-63.

Vos JJ, Walraven PL. On the derivation of the foveal receptor primaries. Vision Res 1971 ;11 :799-818.

Walraven PL. A closer look at the tritanopic convergence point. Vision Res 1974;14:1339-43.

Walsh, D. V., Robinson, J., Jurek, G. M., Capó-Aponte, J. E., Riggs, D. W., & Temme, L. A. (2016). A Performance Comparison of Color Vision Tests for Military Screening. Aerospace Medicine and Human Performance, 87(4), 382-387.

Wright WD. The characteristics of tritanopia. J Opt Soc Am 1952;42:509-20.

Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC medical research methodology, 13*(1), 61.

Wyszecki, G., & Stiles, W. (1982). Colour science: concepts and method.

Zhu, Y., Luo, M. R., Fischer, S., Bodrogi, P., & Khanh, T. Q. (2017). Long-term memory color investigation: culture effect and experimental setting factors. *JOSA A*, *34*(10), 1757-1768.

Appendix 1

Colour Differences in the Normal Space

In order to calculate the colour differences in this space (L*a*b*), the chromaticity

coordinates (x,y,z) for each colour that has been measured were converted to tristimulus

values using the following equations

Y = the luminance   --- (1)

$$X = \frac{Y*x}{y}$$ --- (2)

$$Z = \frac{Y*z}{y}$$ --- (3)

where X, Y, and Z are the tristimulus values.

The tristimulus values for each colour were converted to CIE L*a*b* space using the

following equations (Wyszecki & Stiles; 1982)

L* = 116 (Y/Yn) $^{1/3}$ − 16   --- (4)

a* = 500 [(X/Xn) $^{1/3}$ − (Y/Yn) $^{1/3}$]   --- (5)

b* = 200 [(Y/Yn) $^{1/3}$ - (Z/Zn) $^{1/3}$]   --- (6)

Where Xn, Yn, and Zn are the stimulus values for the reference white. The tristimulus values

of the white diffusing plate were as the reference white for the F-D15, and the white screen

(R=225, G=255, B=255) was the white reference used for the ColorDx D15.

Then, the total colour differences (ΔE) between two caps/disks/rectangles were calculated

using the following equation

$$\Delta E_N^* = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2} \quad \text{---} \quad (7)$$

Where ΔL, Δa, and Δb represent the difference between the two colours coordinates.

Colour Differences in the Dichromatic Space

The tristimulus values were converted first to LMS cone fundamental for 2 degrees using Golz and Macleod (2003) equation 5

$$L = 0.17156X + 0.52901Y - 0.02199Z \quad \text{---} \quad (8)$$

$$M = -0.15955X + 0.48553Y - 0.04298Z \quad \text{---} \quad (9)$$

$$S = 0.01916X - 0.03989Y + 1.03993Z \quad \text{---} \quad (10)$$

The next step was to convert colours into protanopic or deuteranopic cone space using Brettel et al (1997) algorithm. The colour stimuli were represented as vectors in three-dimensional in the LMS space. White stimuli for CVNs assumed to be perceived as neutral for dichromats. The stimuli of the yellow wavelength (575 nm) and the blue wavelength (475 nm) were assumed to be same in deuteranopes and protanopes. The next section is the representation of the algorithm:

The colour stimulus in LMS space is presented as Q for colour normal and as Q' for the dichromate. This colour stimulus can be projected into a plane determined by the monochromatic anchor stimulus A, the stimuli of the white colour E, and origin O. For a given E (L_E, M_E, S_E) and the A (L_A, M_A, S_A), the coordinates $L_{Q'}$, $M_{Q'}$, $S_{Q'}$ for a stimulus Q' can be calculated by

$$aL_{Q'} + bM_{Q'} + cS_{Q'} = 0 \quad \text{---} \quad (8)$$

with

$$a = M_E S_A - S_E M_A \quad \text{-- (9)}$$

$$b = S_E L_A - L_E S_A \quad \text{--- (10)}$$

$$c = L_E M_A - M_E L_A \quad \text{-- (11)}$$

where the L_A, M_A, and S_A are the vector components of the stimulus anchor A and L_E, M_E,

and S_E are the vector components of the white at equal energy (Brettel et al.; 1997).

For a given Q stimulus, the following equation is for transformation into **protanopic** space

$$L_{Q'} = - (bM_Q + cS_Q) /a \quad \text{------------------------------------- (12)}$$

$$M_{Q'} = M_Q \quad \text{------------------------------------- (13)}$$

$$S_{Q'} = S_Q \quad \text{------------------------------------- (14)}$$

If the $S_Q/ M_Q < S_E/ M_E$, then $\lambda_A = 575$ nm; else $\lambda_A = 475$ nm (Brettel et al; 1997)

The next step was to calculate back the tristimulus values (X, Y, and Z) from the LMS

stimuli (L_Q, $M_{Q'}$, $S_{Q'}$) using the inverse of (Golz, MacLeod; 2003) LMS matrix.

$$X = 2.892L - 3.135M + 0.191S \quad \text{--------------------------------------- (15)}$$

$$Y = 0.952L + 1.021M - 0.022S \quad \text{--------------------------------------- (16)}$$

$$Z = -0.017L + 0.097 + 0.957S \quad \text{--------------------------------------- (17)}$$

The tristimulus values for each colour were converted to CIE L*a*b* space using the

equations 4, 5, and 6. The final step is to calculate the ΔE for protanopic cone space using the

following equation

$$\Delta E_p = [(\Delta a)^2 + (\Delta b)^2 + (\Delta c)^2]^{1/2} \qquad \text{-------------------------------------- (18)}$$

where Δa, Δb, and Δc are the colour difference between two different colours in protanopic space.

For a given Q stimulus, the following equations are for transformation into **deuteranopic** space

$$L_{Q'} = L_Q \qquad \text{--- (19)}$$

$$M_{Q'} = - (aL_Q + cS_Q) / b \qquad \text{-- (20)}$$

$$S_{Q'} = S_Q \qquad \text{--- (21)}$$

If the $S_Q / L_Q < S_E / L_E$, then $\lambda_A = 575$ nm; else $\lambda_A = 475$ nm

Next, tristimulus values (X, Y, and Z) was calculated back using equation 15. The tristimulus values for each colour were converted to CIE L*a*b* space using the equations 4, 5, and 6.

The final step is to calculate the ΔE for deuteranopic cone space using the following equation

$$\Delta E_d = [(\Delta a)^2 + (\Delta b)^2 + (\Delta c)^2]^{1/2} \quad \text{---}$$

(22)

where Δa, Δb, and Δc are the colour difference between two different colours in deuteranopic space.

Appendix 2

Table A. The model colour differences values (ΔE) for the F-D15in normal, protan, deutan, and tritan spaces

Cap		REF	1	2	3	4	5	6	7	8	9	10	11	12	13	14
REF	Normal	0.0														
	Protan	0.0														
	Deutan	0.0														
	Tritan	0.0														
1	Normal	11.3	0.0													
	Protan	10.9	0.0													
	Deutan	10.2	0.0													
	Tritan	4.5	0.0													
2	Normal	16.7	5.7	0.0												
	Protan	15.7	4.8	0.0												
	Deutan	14.2	4.0	0.0												
	Tritan	2.6	2.1	0.0												
3	Normal	23.2	12.2	6.6	0.0											
	Protan	21.5	10.6	5.8	0.0											
	Deutan	19.3	9.1	5.1	0.0											
	Tritan	3.1	3.1	1.7	0.0											

4	Normal	28.9	17.7	12.2	5.9	0.0								
	Protan	27.3	16.4	11.6	5.8	0.0								
	Deutan	24.9	14.7	10.7	5.6	0.0								
	Tritan	4.8	2.1	2.4	2.1	0.0								
5	Normal	33.9	22.6	17.3	11.2	5.5	0.0							
	Protan	32.7	21.7	17.0	11.2	5.4	0.0							
	Deutan	30.5	20.3	16.3	11.2	5.7	0.0							
	Tritan	6.3	4.2	4.3	3.2	2.1	0.0							
6	Normal	42.2	31.0	26.0	20.3	14.7	9.3	0.0						
	Protan	41.6	30.6	25.9	20.1	14.3	8.9	0.0						
	Deutan	40.2	29.9	26.0	20.9	15.3	9.7	0.0						
	Tritan	11.1	8.7	9.1	8.0	6.8	4.8	0.0						
7	Normal	50.9	40.1	35.8	30.9	25.6	20.5	11.6	0.0					
	Protan	50.7	39.8	35.1	29.3	23.5	18.2	9.2	0.0					
	Deutan	51.0	40.8	36.8	31.7	26.1	20.5	10.9	0.0					
	Tritan	16.9	13.5	14.6	13.8	12.2	10.7	6.4	0.0					
8	Normal	55.6	46.0	43.0	39.5	35.3	31.0	23.6	13.5	0.0				
	Protan	53.5	42.6	37.9	32.1	26.3	21.1	12.2	3.0	0.0				
	Deutan	57.0	46.8	42.8	37.7	32.2	26.5	16.9	6.1	0.0				
	Tritan	26.9	23.5	24.7	23.9	22.2	20.7	16.1	10.2	0.0				

256

9	Normal	53.7	45.4	43.4	41.3	38.1	34.7	29.2	21.3	9.4	0.0				
	Protan	48.5	37.6	32.9	27.2	21.4	16.3	7.7	3.1	5.1	0.0				
	Deutan	54.1	43.9	40.0	34.8	29.3	23.6	14.0	3.3	2.9	0.0				
	Tritan	32.1	28.6	29.8	29.0	27.4	25.9	21.5	15.7	5.7	0.0				
10	Normal	47.2	40.8	40.3	40.0	38.4	36.6	33.9	29.5	20.5	12.0	0.0			
	Protan	37.7	26.9	22.3	16.7	11.2	7.0	6.0	13.5	16.1	10.9	0.0			
	Deutan	44.9	34.7	30.7	25.6	20.1	14.4	4.8	6.2	12.1	9.2	0.0			
	Tritan	36.3	32.8	34.0	33.3	31.6	30.2	25.9	20.2	10.4	4.6	0.0			
11	Normal	42.2	37.8	38.7	39.9	39.7	39.2	38.7	36.6	29.5	21.4	9.5	0.0		
	Protan	28.7	18.0	13.8	8.8	5.6	7.4	14.3	22.8	25.3	20.2	9.3	0.0		
	Deutan	36.5	26.3	22.3	17.2	11.7	6.0	3.6	14.5	20.5	17.6	8.4	0.0		
	Tritan	36.0	32.6	33.8	33.0	31.4	29.8	25.5	19.8	10.0	4.3	1.3	0.0		
12	Normal	38.5	35.0	36.4	38.4	38.9	39.0	39.6	38.8	33.0	25.5	13.9	4.8	0.0	
	Protan	24.4	14.0	10.2	6.7	7.2	11.2	18.9	27.4	30.0	24.8	13.9	4.6	0.0	
	Deutan	31.7	21.4	17.5	12.4	6.8	1.5	8.5	19.3	25.4	22.5	13.3	5.0	0.0	
	Tritan	34.6	31.1	32.4	31.6	29.9	28.5	24.1	18.4	8.5	2.8	2.0	1.5	0.0	
13	Normal	33.9	32.1	34.5	37.6	39.1	40.1	42.2	43.0	38.6	31.6	20.3	11.4	6.7	0.0
	Protan	17.9	8.0	5.9	7.2	11.6	16.7	25.1	33.8	36.5	31.3	20.4	11.2	6.6	0.0
	Deutan	24.9	14.7	10.7	5.6	0.4	5.7	15.3	26.1	32.1	29.3	20.0	11.7	6.8	0.0
	Tritan	30.4	26.8	28.1	27.3	25.6	24.1	19.7	13.8	4.0	1.9	6.5	6.1	4.6	0.0

14	**Normal**	28.6	28.3	31.5	35.4	37.8	39.7	43.0	45.4	42.6	36.4	25.7	17.2	12.5	6.1	0.0
	Protan	13.0	5.1	6.7	11.1	16.3	21.5	30.1	38.9	41.6	36.5	25.5	16.3	11.7	5.1	0.0
	Deutan	19.0	8.7	4.8	0.6	6.0	11.6	21.2	32.1	38.1	35.2	26.0	17.6	12.7	6.0	0.0
	Tritan	26.4	22.9	24.1	23.3	21.7	20.1	15.6	9.6	0.7	6.4	11.0	10.6	9.2	4.6	0.0
15	**Normal**	23.1	24.2	28.1	32.8	36.0	38.6	43.2	47.1	45.9	40.5	30.5	22.7	18.1	12.1	6.0
	Protan	9.2	5.6	9.4	14.6	20.1	25.4	34.1	43.0	45.7	40.6	29.6	20.4	15.9	9.3	4.1
	Deutan	13.7	3.6	1.3	5.7	11.3	16.9	26.5	37.4	43.4	40.5	31.3	22.9	18.0	11.3	5.3
	Tritan	21.4	18.1	19.1	18.3	16.7	15.1	10.6	4.6	6.0	11.8	16.4	16.0	14.5	9.9	5.4

Table B. The model colour differences values (ΔE) for the surface pro ColorDx D15 in normal, protan, deutan, and tritan spaces

Cap		REF	1	2	3	4	5	6	7	8	9	10	11	12	13	14
REF	Normal	0.0														
	Protan	0.0														
	Deutan	0.0														
	Tritan	0.0														
1	Normal	9.0	0.0													
	Protan	9.0	0.0													
	Deutan	8.8	0.0													
	Tritan	4.1	0.0													
2	Normal	14.6	6.2	0.0												
	Protan	13.9	4.9	0.0												
	Deutan	12.7	3.9	0.0												
	Tritan	3.2	0.9	0.0												
3	Normal	20.4	11.9	5.8	0.0											
	Protan	19.2	10.2	5.3	0.0											
	Deutan	17.4	8.7	4.8	0.0											
	Tritan	3.7	0.6	0.5	0.0											
	Normal	25.6	16.8	11.0	5.4	0.0										
	Protan	24.6	15.6	10.7	5.4	0.0										

4	Deutan	22.8	14.0	10.1	5.3	0.0										
	Tritan	5.7	1.6	2.5	2.0	0.0										
5	Normal	29.3	20.4	14.9	9.6	4.2	0.0									
	Protan	28.6	19.6	14.8	9.5	4.1	0.0									
	Deutan	27.2	18.4	14.5	9.7	4.4	0.0									
	Tritan	7.9	3.8	4.7	4.2	2.2	0.0									
6	Normal	35.9	26.9	22.1	17.3	12.2	8.1	0.0								
	Protan	35.8	26.8	21.9	16.6	11.2	7.2	0.0								
	Deutan	35.3	26.5	22.6	17.8	12.5	8.1	0.0								
	Tritan	12.9	8.8	9.7	9.2	7.2		0.0								
7	Normal	43.8	35.0	31.1	26.8	22.0	18.0	10.0	0.0							
	Protan	43.5	34.5	29.7	24.4	19.0	15.0	7.8	0.0							
	Deutan	44.6	35.8	31.9	27.2	21.8	17.4	9.3	0.0							
	Tritan	19.8	15.7	16.6	16.2	14.2	12.0	7.0	0.0							
8	Normal	50.0	41.9	39.5	36.5	32.5	29.0	21.8	12.8	0.0						
	Protan	46.9	38.0	33.2	27.9	22.6	18.6	11.4	3.8	0.0						
	Deutan	51.0	42.3	38.4	33.6	28.3	23.9	15.8	6.5	0.0						
	Tritan	30.4	26.3	27.2	26.7	24.7	22.5	17.5	10.7	0.0						
	Normal	48.5	41.1	39.7	37.7	34.5	31.5	25.4	17.9	7.0	0.0					
	Protan	42.8	33.8	29.0	23.8	18.6	14.6	7.6	2.5	4.2	0.0					

#	Type															
9	Deutan	48.4	39.6	35.7	31.0	25.7	21.3	13.1	3.8	2.7	0.0					
	Tritan	34.3	30.2	31.1	30.7	28.6	26.5	21.5	14.8	4.4	0.0					
10	Normal	43.7	37.6	37.8	37.4	35.6	33.7	29.7	25.2	17.2	10.7	0.0				
	Protan	33.3	24.4	19.7	14.7	9.7	6.2	4.4	10.7	13.7	9.5	0.0				
	Deutan	40.5	31.8	27.8	23.1	17.8	13.4	5.3	4.2	10.6	7.9	0.0				
	Tritan	37.2	33.1	34.0	33.5	31.5	29.3	24.4	17.9	7.5	3.1	0.0				
11	Normal	40.4	35.3	36.6	37.1	36.2	35.1	32.4	29.6	23.2	16.9	6.3	0.0			
	Protan	27.2	18.4	13.8	9.1	5.3	4.7	9.6	16.8	19.9	15.7	6.2	0.0			
	Deutan	34.9	26.1	22.2	17.5	12.2	7.8	1.1	9.8	16.2	13.5	5.6	0.0			
	Tritan	37.3	33.3	34.2	33.7	31.7	29.5	24.6	18.0	7.7	3.3	0.2	0.0			
12	Normal	37.4	33.2	35.2	36.5	36.4	35.8	34.1	32.6	27.5	21.5	11.1	4.9	0.0		
	Protan	22.4	13.6	9.2	5.3	5.1	7.8	14.1	21.6	24.8	20.6	11.1	4.9	0.0		
	Deutan	30.1	21.3	17.4	12.7	7.4	3.2	5.3	14.6	21.0	18.3	10.4	4.8	0.0		
	Tritan	36.3	32.2	33.1	32.7	30.7	28.5	23.6	17.0	6.6	2.2	0.9	1.1	0.0		
13	Normal	32.4	29.1	31.9	34.0	34.7	34.8	34.4	34.6	31.4	26.0	16.3	10.4	5.8	0.0	
	Protan	17.1	8.4	4.7	4.5	8.6	12.3	19.1	26.7	30.0	25.8	16.3	10.1	5.2	0.0	
	Deutan	24.2	15.4	11.5	6.8	1.9	3.3	11.2	20.5	26.9	24.3	16.4	10.7	5.9	0.0	
	Tritan	32.7	28.6	29.5	29.1	27.0	24.8	19.9	13.1	2.7	1.9	4.9	5.1	4.0	0.0	
	Normal	27.7	25.9	29.8	32.9	34.6	35.4	36.5	38.2	36.6	31.7	22.5	16.7	12.2	6.4	0.0
	Protan	11.3	3.5	4.6	9.2	14.2	18.1	25.1	32.7	36.0	31.8	22.3	16.1	11.2	6.0	0.0

14	Deutan	17.7	9.0	5.1	1.0	5.2	9.5	17.6	26.9	33.4	30.7	22.8	17.2	12.4	6.5	0.0
	Tritan	28.9	24.8	25.7	25.3	23.2	21.0	16.0	9.1	1.8	6.1	9.2	9.4	8.3	4.3	0.0
15	Normal	21.9	21.3	26.0	29.9	32.4	33.9	36.4	39.7	40.0	35.8	27.4	22.2	17.9	12.2	6.2
	Protan	7.4	3.4	7.7	12.8	18.0	22.0	29.0	36.6	40.0	35.8	26.3	20.1	15.2	10.0	4.0
	Deutan	12.4	3.7	0.6	5.0	10.4	14.7	22.9	32.2	38.6	36.0	28.1	22.5	17.7	11.8	5.3
	Tritan	22.8	18.7	19.6	19.2	17.1	15.0	10.0	3.0	7.8	12.1	15.2	15.4	14.3	10.4	6.2

Appendix 3

Table A the questionnaire that has been used in this study.

	Question	Yes	No
1	Are you being treated or do you have Glaucoma; Optic Neuritis; Multiple Sclerosis; Diabetes		
2	Other than wearing spectacles or contact lenses and/ or having colour vision problem, are you aware of any other vision problems?		
4	Do you use any of these medications? (circle) Chloroquine, Hydroxychloroquine "Plaquenil" or Digitalis "Digoxin"		
5	Have you had a cataract surgery?		
6	Do you have a colour vision defect? (if 6 answer is NO please proceed to 7)		
6a	At which age did you first become aware of the problem with colour judgments? Age [ ]		
6b	Who informed you that you have a colour vision defect? ……………………………………………..		
6c	Do you have any difficulty with colour judgments in your daily activities? If yes, what is the major problem		
6d	Did your colour vision problem affect your career choice?		

7	Do you have any difficulty identifying the colour of traffic lights?		
8	Do you have any problem finding traffic signal lights at night when there are numerous other street lights surrounding the traffic signal or in the background?		
9	At some intersections, there are cross walk signals for pedestrians. The hand symbol is a reddish orange and indicates that is it unsafe to cross. What colour is the man-figure which indicates that it is safe to cross? _____		

Appendix 4

Table A. The ROC area, the cut-off point, AC1, PreP, and PreF values. CI is the confidence interval.

Test	CN lantern Distance (m)	ROC		Cut-off point	AC1		PreP		PreF	
		Area	95% CI		value	95% CI	value	95% CI	value	95% CI
Farnsworth D15 (Number of Crossings)	4.6	0.77	0.60 to 0.95	3.167	0.43	0.14 to 0.68	0.12	0.04 to 0.30	1	0.90 to 1
	2.3	0.83	0.73 to 0.93	3.167	0.49	0.25 to 0.72	0.24	0.11 to 0.43	1	0.9 to 1
	1.15	0.71	0.56 to 0.85	3.833	0.48	0.25 to 0.71	0.65	0.46 to 0.82	0.79	0.63 to 0.90
	0.57	0.82	0.71 to 0.93	3.833	0.5	0.28 to 0.72	0.88	0.71 to 0.96	0.65	0.48 to 0.79
Farnsworth D15 (C-index)	4.6	0.81	0.62 to 1.00	2.017	0.46	0.22 to 0.70	0.12	0.04 to 0.31	1	0.90 to 1
	2.3	0.85	0.74 to 0.96	2.017	0.52	0.29 to 0.75	0.25	0.12 to 0.45	1	0.9 to 1
	1.15	0.69	0.54 to 0.85	2.483	0.48	0.25 to 0.71	0.65	0.46 to 0.81	0.76	0.6 to 0.88
	0.57	0.79	0.67 to 0.91	2.483	0.5	0.28 to 0.72	0.88	0.71 to 0.96	0.65	0.48 to 0.79
ColorDx D15 (Number of Crossings)	4.6	0.73	0.54 to 0.92	4	0.39	0.14 to 0.65	0.12	0.04 to 0.29	1	0.90 to 1
	2.3	0.83	0.72 to 0.93	4	0.45	0.21 to 0.70	0.23	0.11 to 0.42	1	0.9 to 1
	1.15	0.73	0.59 to 87	4	0.48	0.25 to 0.71	0.66	0.46 to 0.80	0.79	0.63 to 0.9
	0.57	0.78	0.65 to 0.9	5.5	0.57	0.36 to 0.78	0.87	0.70 to 0.95	0.7	0.52 to 0.83
	4.6	0.73	0.54 to 0.92	2.322	0.39	0.14 to 0.65	0.12	0.04 to 0.29	1	0.90 to 1
	2.3	0.83	0.71 to 0.95	2.322	0.45	0.21 to 0.70	0.23	0.11 to 0.42	1	0.9 to 1

ColorDx D15 (C-index)	1.15	0.75	0.61 to 0.89	2.678	0.48	0.25 to 0.71	0.64	0.46 to 0.79	0.81	0.65 to 0.91
	0.57	0.79	0.66 to 0.91	2.678	0.57	0.36 to 0.78	0.89	0.73 to 0.96	0.69	0.51 to 0.82
HRR (Diagnostic Plates)	4.6	0.69	0.43 to 0.96							
	2.3	0.85	0.68 to 1	Moderate or severe	0.87	0.76 to 0.98	0.5	0.22 to 0.79	0.96	0.9 to 0.99
	1.15	0.76	0.63 to 0.89	Moderate or severe	0.56	0.37 to 0.8	1	0.68 to 1	0.71	0.57 to 0.81
	0.57	0.68	0.55 to 0.82	Moderate or sever	0.186	0 to 0.46	1	0.68 to 1	0.49	0.36 to 0.62
ColorDx (Diagnostic Plates)	4.6	0.75	0.43 to 1							
	2.3	0.78	0.55 to 1	Moderate or severe	0.82	0.69 to 0.95	0.4	0.17 to 0.69	0.96	0.86 to 0.99
	1.15	0.72	0.57 to 0.86	Moderate or severe	0.63	0.43 to 0.84	1	0.72 to 1	0.74	0.6 to 0.84
	0.57	0.65	0.51 to 0.79	Moderate or severe	0.24	0 to 0.51	1	0.72 to 1	0.51	0.38 to 0.64
CAD	4.6	0.62	0.40 to 0.84							
	2.3	0.79	0.61 to 0.96	14.47	0.71	0.53 to 0.88	0.31	0.14 to 0.56	0.98	0.88 to 1
	1.15	0.74	0.60 to 0.88	22.45	0.47	0.23 to 0.70	0.62	0.44 to 0.77	0.83	0.67 to 0.93
	0.57	0.69	0.55 to 0.83	23	0.36	0.12 to 0.60	0.75	0.58 to 0.88	0.6	0.42 to 0.75
CCT (Elliptical Area)	4.6	0.53	0.26 to 0.8							
	2.3	0.58	0.32 to 0.83							
	1.15	0.63	0.48 to 0.77							
	0.57	0.7	0.56 to 0.84	$1.12*10-3$	0.41	0.17 to 0.65	0.76	0.59 to 0.87	0.63	0.44 to 0.79

CCT (Ellipse Length)	4.6	0.49	0.27 to 0.71						
	2.3	0.6	0.34 to 0.85						
	1.15	0.62	0.47 to 0.76						
	0.57	0.69	0.54 to 0.83	0.0984	0.44	0.21 to 0.68	0.78	0.61 to 0.89	0.64	0.46 to 0.79
LandC (Average)	4.6	0.41	0.23 to 0.59						
	2.3	0.56	0.37 to 0.75						
	1.15	0.53	0.38 to 0.68						
	0.57	0.44	0.29 to 0.59						
LandC (Maximum)	4.6	0.46	0.32 to 0.6						
	2.3	0.63	0.46 to 0.8						
	1.15	0.52	0.37 to 0.67						
	0.57	0.46	0.31 to 0.6						
RCCT (Average)	4.6	0.5	0.04 to 0.96						
	2.3	0.41	0.14 to 0.69						
	1.15	0.54	0.39 to 0.70						
	0.57	0.55	0.40 to 0.70						
RCCT (Minimum)	4.6	0.47	0.06 to 0.89						
	2.3	0.59	0.35 to 0.83						
	1.15	0.56	0.41 to 0.71						
	0.57	0.55	0.40 to 0.70						

Poster 1

Modified Image from the

Canadian Armed Forces Image

Gallery.

ww.combatcamera.forces.gc.ca

Colour Vision Tests: The Next Generation

We are determining whether new colour vision tests for testing pilots and aircrew are as good as the current tests.

We looking for participants who have **NORMAL COLOUR VISION**, **COLOUR VISION DEFICIENCIES,** who **SUSPECT** THAT THEY HAVE A **COLOUR VISION PROBLEM,** and individuals who are **CURIOUS** about their **COLOUR VISION**.

There will be two sessions separated by approximately 2 weeks. In each session, your colour vision will assessed with a variety of different tests which determine how well you see shades of colour. Each session takes about 2 hours to complete.

After participating in the sessions, you will receive $45

If you are between17 and 60 yrs and you would like more information or are interested in participating, please contact us at

> **Colourvision.study@uwaterloo.ca**
> School of Optometry and Vision Science
> University of Waterloo
> **226-789-7679**
> This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee. ORE #: 20996

Colourvision.study@uwaterloo.ca

Poster 2

Modified Image from the Canadian Armed Forces Image Gallery.

www.combatcamera.forces.gc.ca

Colour Vision Tests: The Next Generation

We are determining whether new colour vision tests for testing pilots and aircrew are as good as the current tests.

We looking for participants who have **NORMAL COLOUR VISION**, **COLOUR VISION DEFICIENCIES,** who **SUSPECT** THAT THEY HAVE A **COLOUR VISION PROBLEM,** and individuals who are **CURIOUS** about their **COLOUR VISION**.

There will be two sessions separated by approximately 2 weeks. In each session, your colour vision will assessed with a variety of different tests which determine how well you see shades of colour. Each session takes about 2 hours to complete.

After participating in the sessions, you will receive $45

If you are between17 and 60 yrs and you would like more information or are interested in participating, please contact us at

Colourvision.study@uwaterloo.ca
School of Optometry and Vision Science
University of Waterloo
226-789-7679

Colourvision.study@uwaterloo.ca

This study has been reviewed and received ethics clearance through a University of Waterloo

Research Ethics Committee. ORE #: 20996

Do you think that you have difficulty distinguishing between colours or are unsure, then this study may be of interest to you

If you are between17 and 60 yrs and you would like more information or are interested in participating, please contact us at
Colourvision.study@uwaterloo.ca
School of Optometry and Vision Science
University of Waterloo
226-789-7679

We are determining the suitability and repeatability of the **next generation of colour vision tests** to be used in Railway.

After participating in two sessions, you will receive **$30**

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee. ORE 21094