Computerized and Non-Computerized Colour Vision Tests

by

Ali Almustanyir

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

Waterloo, Ontario, Canada, 2014

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this 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.

Abstract

Introduction: Measuring colour discrimination of people who carry out tasks where colour is used to convey information and accurate colour judgments are essential for safe and efficient performance of the task is important in order to ensure that they can carry out the tasks. Individuals with congenital colour vision deficiencies are at a greater risk of making an error in colour judgment and this is the primary reason for colour vision testing in industry. Today, there is a large number of colour vision tests to detect colour vision deficiencies and/or estimate one's ability to discriminate colours.

Purpose: The purpose of this study is to determine the validity and repeatability of a new colour vision test "Colour Vision Reaction Time" (CVRT) for screening for colour vision defect. The study will also determine the repeatability of a selection of clinical colour vision tests, which are currently in use.

Material and methods: The test series was administered to 75 colour normal subjects and 47 participants with red-green defects. Colour vision was classified based on the Nagel anomaloscope. In the pseudoisochromatic tests (the Hardy, Rand, Rittler 4th edition (HRR), Ishihara 38 plate edition and Pseudoisochromatic Plates Ishihara Compatible (PIPC) tests) subjects are required to identify the coloured figure within a background of a different colour. For the Colour Vision Reaction Time (CVRT) test, subjects need to locate a coloured Square on a computer screen using a mouse. The Cone Contrast Sensitivity Test (CCST) requires individuals to identify coloured letters that may appear in a gray background on the computer screen.   A prototype of the ColorDx (pColorDx) test is similar to the printed pseudoisochromatic plates except that the plates are displayed on a computer screen. The

Farnsworth-Munsell D15 (D15) test requires subjects to arrange coloured caps in order according to colour starting from a fixed cap.

Results: The agreement of the printed pseudoisochromatic tests with the anomaloscope in terms of screening for red-green defects was good with kappa (κ) coefficient of agreement value of 0.96 or more on all three tests. The repeatability of the three tests was good with κ coefficient of 0.96 or more on the three tests. Both HRR and PIPC tests can screen for blue-yellow defects. There were 2 deuteranomalous subjects at the first visit and a different deuteranomalous individual at the second visit who made a single blue-yellow error in the HRR test. In the PIPC test, only one deuteranomalous subject failed the blue-yellow screening plates at the first visit with two errors. In terms of the classification as either protan or deutan, the agreement with the Nagel anomaloscope was perfect with the HRR test and acceptable with the Ishihara, but only fair for the PIPC test. The agreement of the repeatability of the classification was perfect with the HRR test and good with the Ishihara test whereas it was reasonable with the PIPC test. The HRR test was designed to classify the severity of the defect and there was a reasonable correlation between the HRR severity and the Nagel anomaloscope matching ranges.

The agreement of the three computerized colour vision tests with the anomaloscope was good with κ coefficients ≥ 0.91. The repeatability of these three tests was good with κ coefficients ≥ 0.98. All the three tests can screen for blue-yellow defects. In the CVRT test, the response times of most subjects to the blue-yellow test figures were within 1.0 standard deviation of the white control value. The single exception was a deuteranomalous subject who did not fail any other blue-yellow screening test. Ten subjects failed the pColorDx

blue-yellow test, whereas 3 subjects failed the CCST S-cone portion. The CCST and pColorDx computer test can classify the defect as protan or deutan. Both tests showed a good-to-perfect agreement with the anomaloscope. The pColorDx test can grade the severity of the defect in terms of mild, moderate to severe. The Spearman rank correlation coefficient with the Nagel matching ranges was only moderate.

The Farnsworth D15 test was included to determine whether there was a difference in the pass rate using the results from the first trial or requiring the subjects to pass on 2 of 3 trials. There was a marginal improvement in the pass rate using the 2 out of 3 rule. The repeatability of the 2 out of 3 trials in the D 15 test showed that there was a good agreement between sessions with κ coefficient of 0.87. In terms of classifying the defect as protan or deutan, based on the visual inspection, there was a good agreement with κ coefficient = 0.83. However, based on the Colour Difference Vector (CDV) angle parameter, all the colour defective subjects were correctly classified. The repeatability of classifying the type of the defect based on the CDV showed perfect agreement between the first and second visit. The D15 can classify the defect as mild versus moderate-to-severe. As expected, the majority of individuals who failed the D15 were classified as having moderate to severe classification on the HRR and pColorDx tests.

Discussion and Conclusion: The current study confirms that the three pseudoisochromatic tests are effective in screening for red-green colour vision defect. The HRR test may be preferred over the Ishihara and PIPC because the sensitivity was marginally higher than the other two tests. Agreement of the diagnostic plates with the Nagel anomaloscope as to whether the colour vision defect was protan or deutan varied across tests. The results from

this study agreed with Birch's (1997) results for the Ishihara in that approximately 85% of the colour defectives were classified correctly as either protan or deutan. However, HRR classification results were slightly better than Cole, et al's. In terms of the severity, our results were similar to Cole et al in that there were a reasonable correlation between the HRR severity and the Nagel anomaloscope matching ranges.

The three computerized colour vision tests are effective in terms of screening for red-green defects. The CCST had the highest agreement with the anomaloscope, but it was not significantly different from the other two tests. However, the pColorDx's ability to grade the severity was moderate, but it was slightly lower than the HRR plates. All three tests are capable of screening tritan defects. Our results suggest that a small number of deutans are likely to fail this portion of these tests.

The D 15 test showed a reasonable repeatability in terms of pass/fail when we used the 2 out of 3 rule and marginally better than performing only one trial on separate days. In terms of the repeatability of classification, the study showed that there was a good agreement between sessions based on the visual inspection and perfect agreement between sessions based on Colour Difference Vector parameters.

Acknowledgements

Dedication

I would like to dedicate my thesis to my family and all of my friends. Special thanks to my parents, my brothers and sisters, and my wife for their 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

x

List of Figures

List of Tables

Chapter 1
Introduction

1.1 General introduction

The number of colour vision tests has increased recently and so it becomes important to know which ones are the most convenient, valid, and reliable. It is not enough to know how the tests can be administered. One must also have an understanding of the test design in order to have high confidence in interpreting an individual's result (Lakowski, 1969a). In addition, it is important to know the purpose of the colour vision test. It can be classified into three levels of needs:

- To screen for acquired or congenital colour vision defect (Dain, 2004). If a given test can divide accurately and quickly subjects into defective colour vision and normal colour vision, then this test is a good screening test.

- To determine the type and severity of colour vision defect.

- To determine whether they have adequate colour vision to carry out a specific occupational task. Individuals with congenital colour vision deficiencies are at a greater risk in making an error in colour judgment and this is the primary reason for colour vision testing at the occupations level (Dain, 2004).

    Colour vision deficiencies are characterized by the inability to distinguish characteristic sets of colours. The severity of the discrimination losses can range from mild to severe depending on the nature of the underlying problem. Colour vision deficiencies can be divided into congenital or acquired (Shin, et al, 2007). The primary difference that distinguishes congenital from acquired is that in a congenital defect the visual system is

1

otherwise normal except for the loss in colour discrimination and the defect remains stable throughout life. If there is an ocular disease or disorder, then it is coincidental to the colour vision defect. In contrast, the acquired colour vision defects are always due to an underlying disease or disorder and some other aspect of visual function is affected by the condition. Furthermore, the colour vision defect can regress and progress along with the underlying condition (Pokorny, et al, 1979).

Congenital colour vision defects are the most common colour vision deficiency in patients under 60 years old (Schneck, et al, 2014). Congenital colour vision deficiencies are classified as either red-green or blue-yellow based on which colours they are more likely to confuse. Individuals with red-green defect confuse colours along the red-green colour axis (red, orange, yellow and green), whereas individuals with blue-yellow defect confuse colours along blue-yellow colour axis (violet, gray and yellow-green). Within these two broad categories, the defect can be divided into dichromatic and anomalous trichromatic based on the number of primaries required to match a coloured light. Dichromatic 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 normal population. On average, dichromats have worse colour discrimination than anomalous trichromats (Pokorny, et al, 1979).

The red-green colour defectives can be further classified based on whether the M-cone or L-cone photopigment is missing or different from the colour normal population. Deuteranomalous have an altered M-cone photopigment and deuteranopes are missing the M-cone photopigment. Protanomalous have an altered L-cone photopigment and protanopes

are missing the L-cone photopigment. Most of the colour confusions for both protanomalous and protanopic individuals are qualitatively similar to the colour confusions for deuteranomalous and deuteranopic individuals (Pokorny et al, 1979).

Congenital blue-yellow defects are very rare. Blue-yellow defects include the dichromatic form (tritanopia) and anomalous trichromats form (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).

Acquired colour vision deficiencies are less common in the younger population, but are very common in the elderly (Schneck et al., 2014). This is not surprising since visual disorders also increase with age. Glaucoma, cataract, macular degeneration, and diabetes are the most common causes of acquired colour vision defects (Kaiser & Boynton, 1996). There are three types of acquired colour vision defects. The most common type is a Type III acquired blue –yellow defect. This type of defect is defined by discrimination losses along the blue-yellow axis and is observed in macular degeneration, glaucoma, diabetes, nuclear cataract, and some optic nerve disorders (Kaiser & Boynton, 1996). There are two types of red-green acquired colour vision defects. Type I acquired red-green defect occurs in photoreceptor/retinal pigment epithelium diseases that become manifest in early life. Individuals with this type of defect tend to have protan settings (i.e. require more red in a mixture with green to match a yellowish standard light) on the anomaloscope along with a loss in colour discrimination along the red-green axis. Type II acquired red-green also shows a decline in hue discrimination along the red-green, but this defect tends to occur with optic nerve diseases such as optic atrophies and optic neuritis (Kaiser & Boynton, 1996).

3

Observers with this type of defect tend to have deutan colour settings (i.e. require more green in a mixture with red to match a yellowish standard light) on the anomaloscope. Table 1 present the comparison between acquired colour vision defects and the congenital defects.

Table 1. Comparing the acquired colour vision defects with the congenital defects

	Acquired	Congenital
Tests result	Usually Unstable	Stable
Colour naming	Good or Unpredictable Colour Confusion	Predictable Confusions
Variation	Varies between eyes/parts of the retina	Bilateral and equal
Axis of confusion	Usually mixed	Clearly defined

1.2 Changes in colour vision with age

Colour discrimination changes throughout life. There is an improvement in colour discrimination performance in a variety of colour vision tests from childhood to adolescence (Ling & Dain, 2008). This improvement is often believed to be due to the ability of the person to understand the test. Colour discrimination remains stable until approximately 40 years and then begins to decline for the reminder of the person's life. Tiffin and Kuhn (1942) assessed colour vision deficiency of 7000 industry worker (using an ad hoc colour identification test) and found that 74% of workers age between 20-25 passed the test, whereas only 32% of those over 55 years old pass the test. The changes in colour vision due to normal aging are due to pupil miosis, which decreases retinal illumination, yellowing of the human lens, and an increase in diseases and disorders that occur in later life. The net

result of these three factors is an increase in the prevalence of tritanomalous- like defects
(Schneck et al, 2014).

1.3 Basic elements of colour vision tests

The basic elements of colour vision tests can be defined as:

1.3.1 Mode of appearance

Mode of appearance determines whether the colours stimuli are presented as surface colours
or projected lights. The Ishihara is an example of surface mode of appearance whereas the
anomaloscope is an example of projected light (Lakowski, 1969a).

1.3.2 Visual task complexity

Visual task complexity is synonymous with the cognitive demand of the test. Some of the
tests are easy and some of them are difficult in terms of their cognitive demands. Most
common tests use numerals or recognizable shapes while other tests require the patient to
arrange the test objects in order of hue. Although, there may be some disagreement, tests that
require ordering by hue are considered to have a low complexity whereas plate tests in which
a figure has to be identified in the presence of luminance and saturation differences have high
complexity. The background and figure colours of the pseudo-isochromatic plates (PIC) each
vary in saturation and brightness. This variation makes it difficult to perceive the number
because the person has to separate the number from the background based on hues that look
more similar to each other. One example of the visual complexity is the vanishing plate in
Figure 4. The figure is a combination of blue-green and green colours within an orange,
yellow and red background. The colour-normal should group the two green colours together

to form the number (5) because these two colours are more similar to each other relative to the background colours.

1.3.3 Stimulus complexity

The number of colours in the test defines the complexity of stimulus. In PIC tests, each plate contains many different coloured dots that are used to produce the figure and background. In contrast, the anomaloscope displays only three lights, which is a simpler stimulus.

1.3.4 Angular subtence

It is important to consider the retinal angular subtense because colour vision is affected by the angular size. Small angular sized stimuli(less than 0.5 degree) will produce tritan defects in people with normal colour vision (Lakowski, 1969a). The defect is due to the lower density of S-cones within the very central region of the fovea.

1.4 Colour confusion and pseudo-isochromatic tests (PIC)

The CIE diagram is the most common technique to show the colour confusion of colour vision defects (Birch, 1993). The straight lines in Figure 1, 2, and 3 are the confusion loci for the protanope, deuteranope, tritanope respectively. Any colours lying on the same line (or near the same line) will appear identical to a person with the corresponding defect if the colours are all the same brightness. For example, in Figure 1, green and yellow will appear identical for protan patient because they fall among the same confusion line. The line of confusion that starts at the spectrum locus and passes through the gray region over to the other side of the diagram defines the neutral axis. This axis also defines the major confusion axis. Protans and deutans have discrimination losses along the red-green axis because the

6

neutral axis is between the red and green colours of the diagram. Tritans have discrimination losses along the blue-yellow axis because the neutral axis lies between the yellow and blue colour of the diagram.

PIC tests are an example of colour vision test that use confusion lines to divide people into colour normal and defective and it called pseudo-isochromatic because the colour normal can distinguish between each colour in the plates and read the figures. The figures are isochromatic to the background for colour defectives if the colours lie on the same line of confusion (Lakowski, 1966). Generally, PIC tests consist of numbers of plates in which dots of various sizes and colours comprise the figure and background.
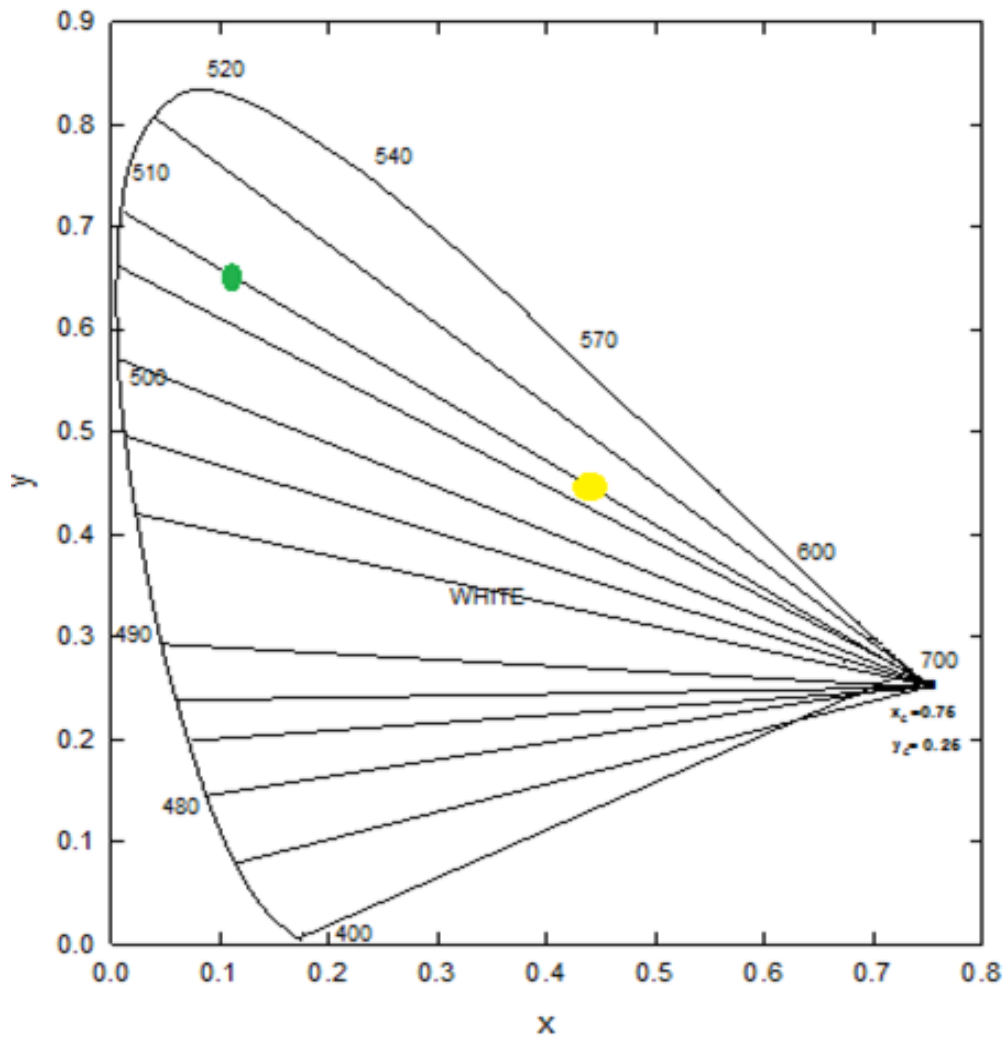
Figure 1. Lines of confusion for a protanope. Green and yellow dots fall in the same line of confusion (Birch, 1993)
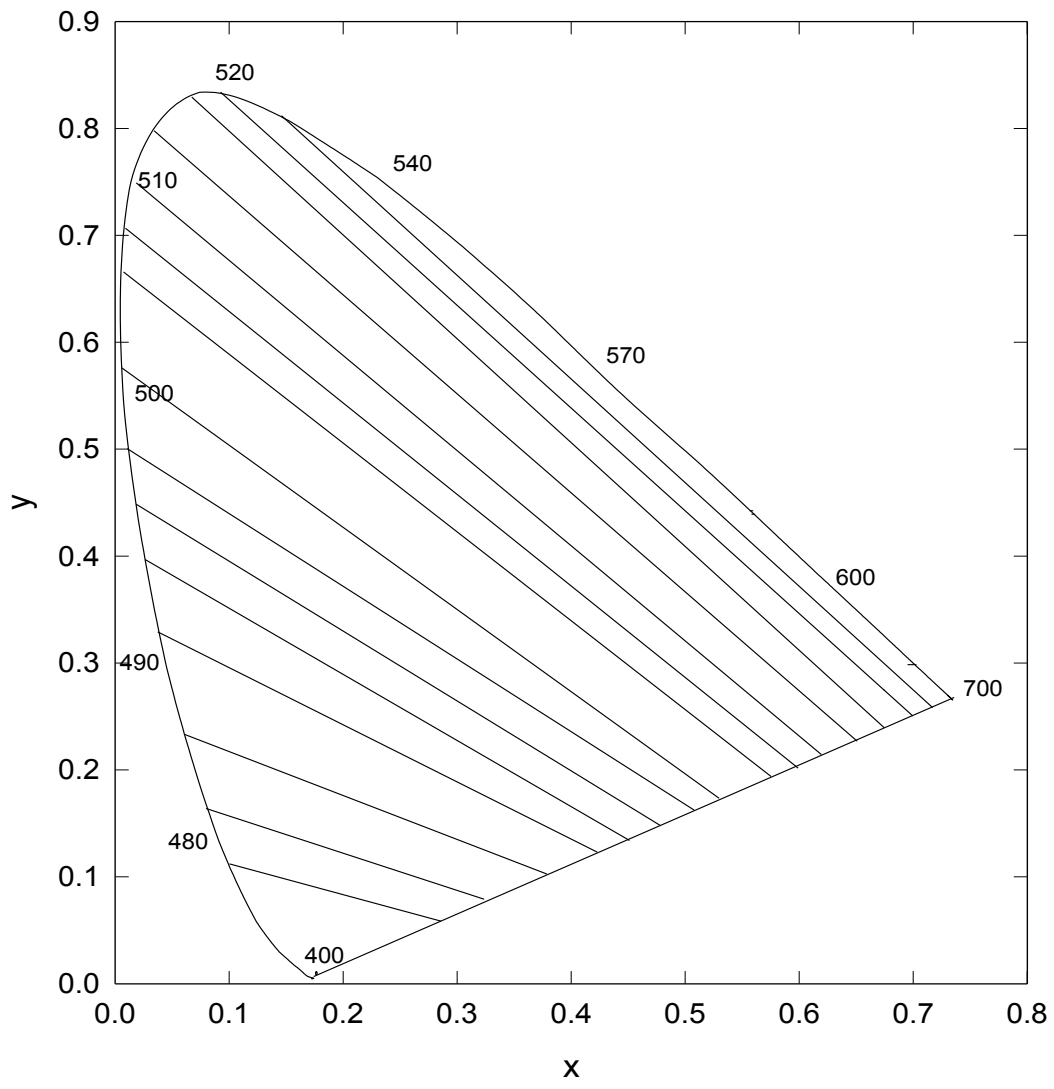
Figure 2:Lines of confusion for a deuteranope (Birch, 1993).
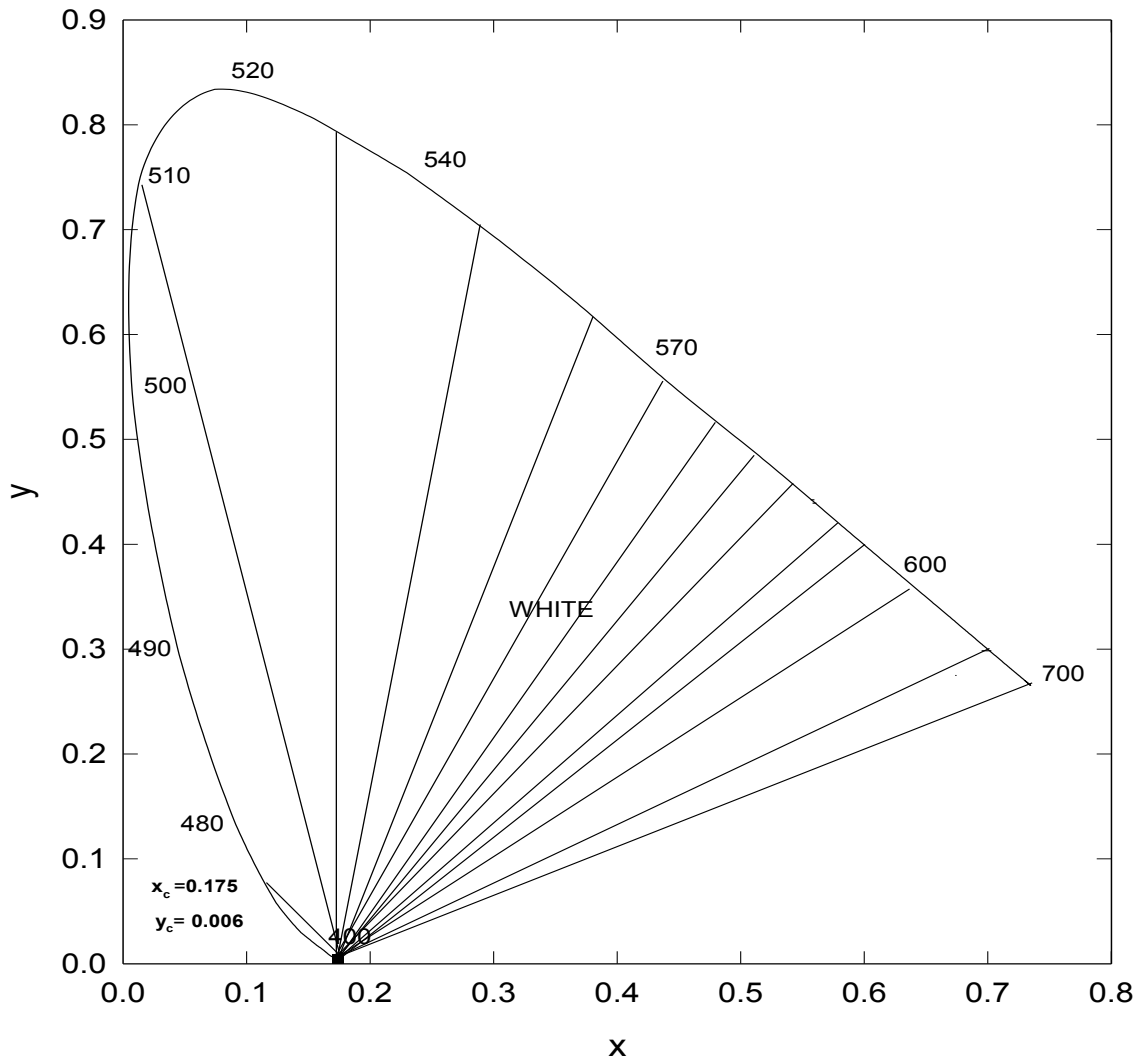
Figure 3. Lines of confusion for a tritanope (Birch, 1993).

1.5 Essential structure of PIC tests

The PIC plates are divided into four designed categories: vanishing, diagnostic,

transformation, and hidden.

1.5.1 Vanishing plates

Vanishing designs are the simplest and most widely used plates. The figure and background colours are chosen to fall along a particular confusion line for a specific dichromat or close to the confusion lines of two or more different dichromats (Shin et al., 2007). Figure 4 shows an example of a vanishing plate in which blue-green and green dots form the figure and orange, yellow and red comprise the background. These particular colours are close to the spectrum locus where the protanopic and deuteranopic lines of confusion are nearly identical so that plate will screen for both protan and deutan defects.
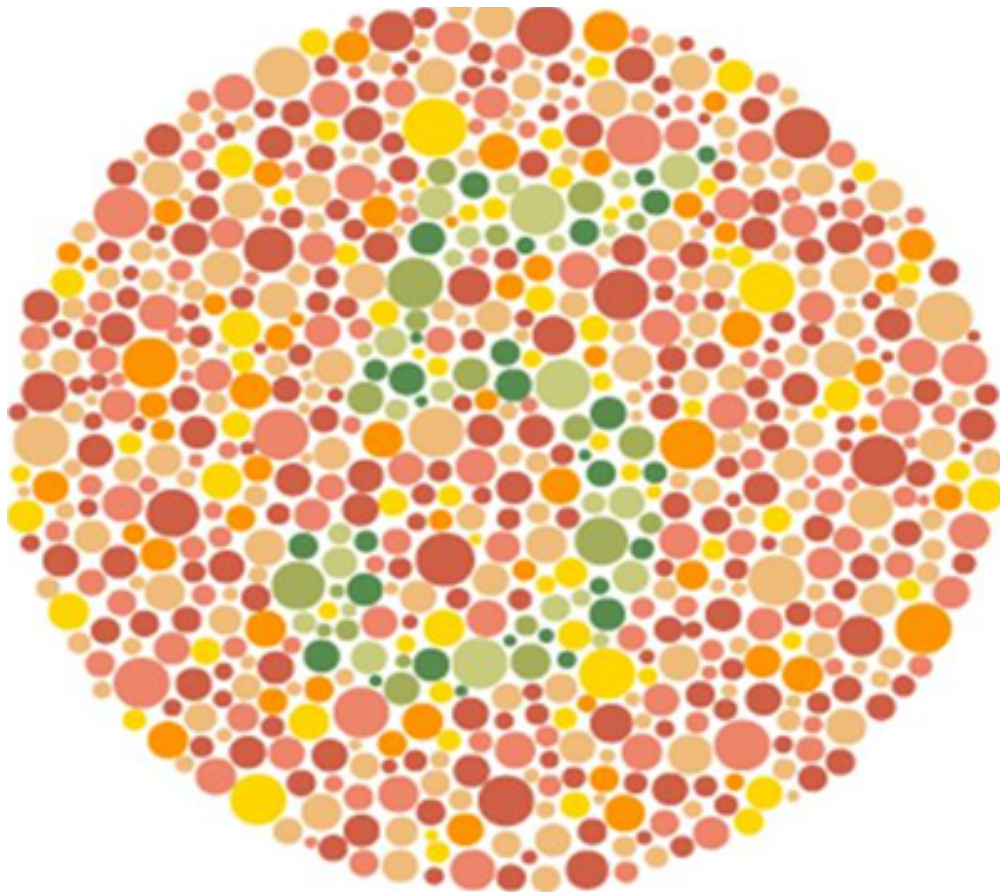
Figure 4. Example of a vanishing plate. (Supplied by Dr T. Waggoner)

Figure 5 shows examples of colorimetric data for the vanishing plate design used for screening and the diagnostic design. For the vanishing plate, the figure and background colours were chosen to fall on the confusion line for both deutans and protans.
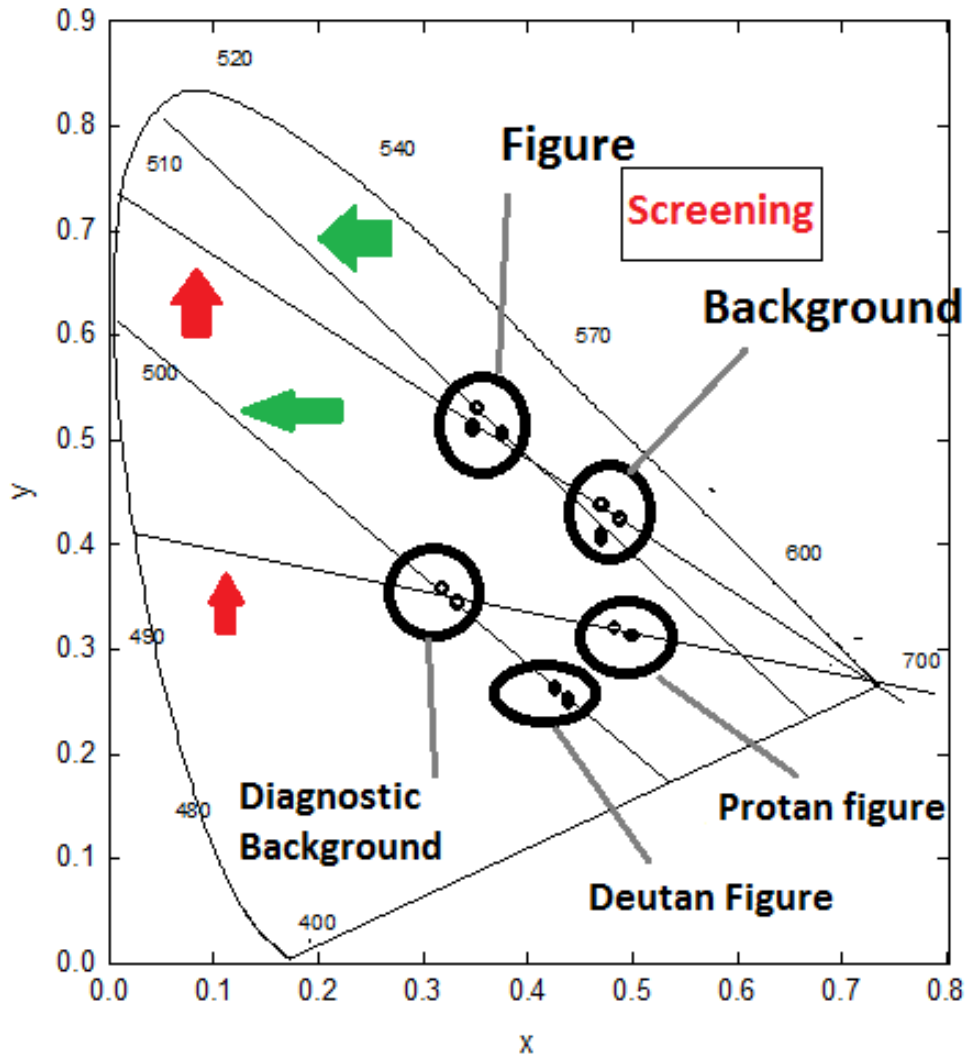
Figure 5. Chromaticities for vanishing and diagnostic plate design. Green arrows indicate deutan line of confusion and red arrows indicate protan line of confusion based on Lakowski 1969b.

1.5.2 Qualitatively diagnostic plates

These are an extension of vanishing plates except two different coloured dots clusters are used for the figures displayed in a common background. The background is usually gray dots. One figure colour is used to identify protans and other figure colour is used to identify deutans. Figure 6 shows an example of plate 15 in the HRR colour vision test. This plate is designed to distinguish between protan and deutan by presenting a pink 3 which is missed by protans and purple 5 which missed by deutans. The pink figure colours fall on the protan confusion line with the gray background and the purple figure colours aligned with gray on the deutan confusion line. In addition, the chroma level of the figure colours can change systemically in order to grade the severity of the defect. The Hardy Rand and Rittler (HRR) is one example. In the HRR diagnostic plates, subjects would be diagnosed as having a severe defect if they make errors, which include the plates with the most saturated colours.
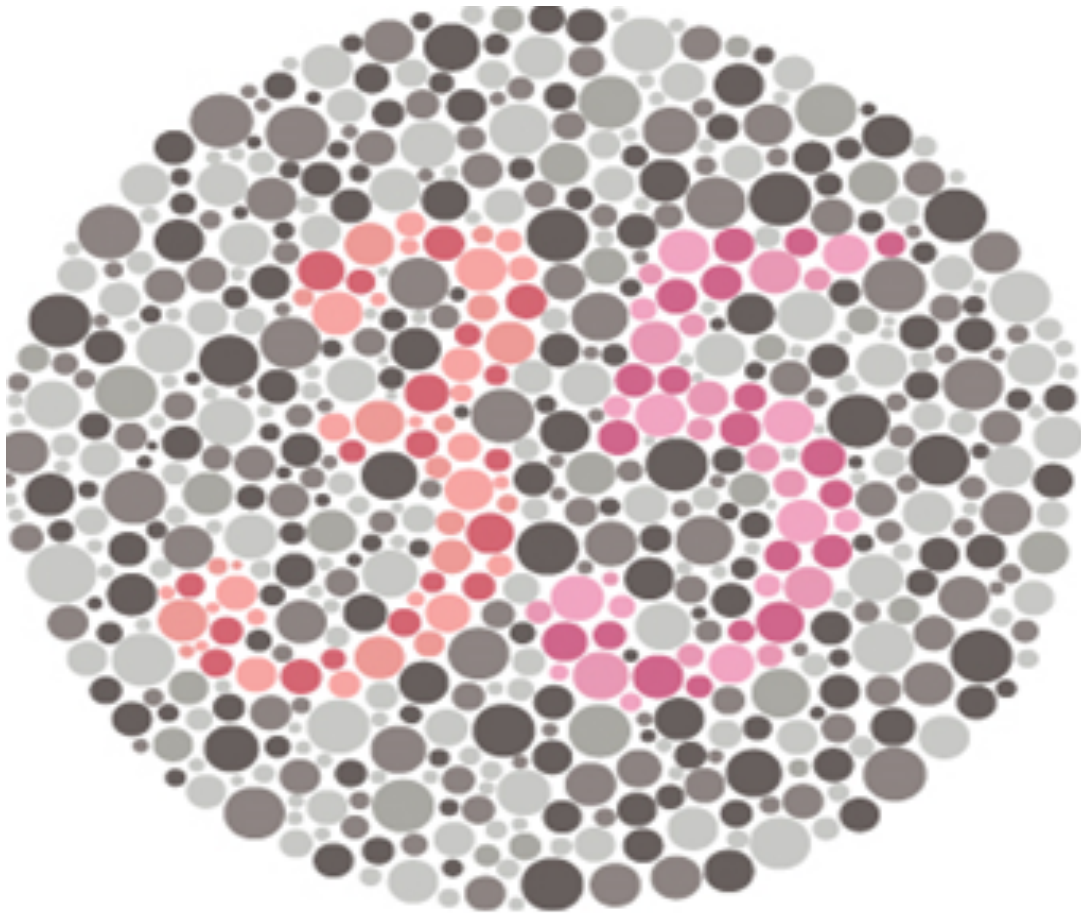
Figure 6. Example of a diagnostic plate. The pink figure is missed by protans and the purple figure is missed by deutans (Supplied by Dr T. Waggoner).

1.5.3 Transformation plates

These are the more interesting designed plates for both the colour-normal and defective

because both will see a number or figure (Lakowski, 1969b). Figure 7 is an example of a

transformation plate in which the blue-green and green colours dots form the figure within

the orange and purple colours dots background for the colour-normal. The two green hues

are more similar to each other than the greens are to the orange background colours, so they

are grouped together to form the number for the colour-normal. For the colour defective, the

green colour compromising part of the 4 is on the same line of confusion as the orange background, but the blue-green portion is not. The green appears identical to the background colour and so the colour-defective perceives just a "1". Similarly, the green in the "5" would appear similar to the orange background, whereas the blue-green does not. However, if you look closely at the figure, there is a purple line of colours running diagonally from the top right of the 5. This purple is on the same line of confusion as the blue-green colours, but not on the same line of confusion with orange background. The colour defective subject should group the blue-green and purple colours together and likely report a 7 instead of a 5.
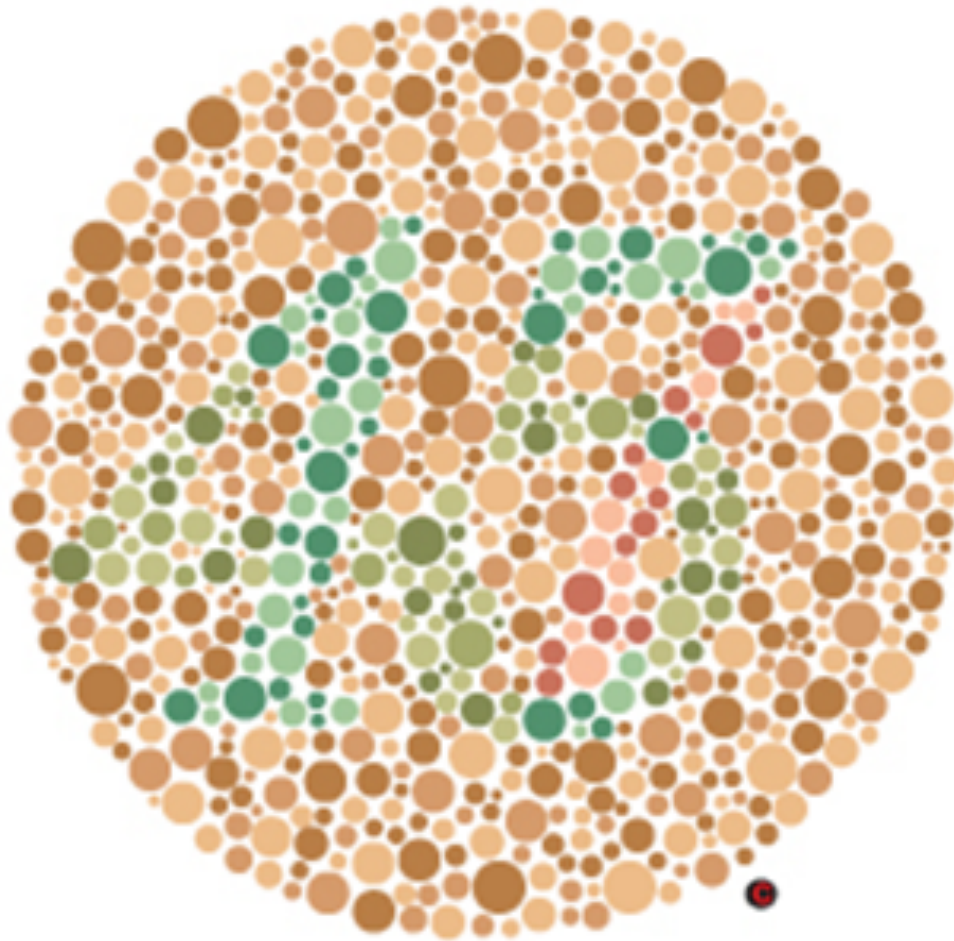
Figure 7. Example of a transformation plate. (Supplied by Dr T. Waggoner)

1.5.4 Hidden digit plates

These plates designed so that the colour defect will see a figure while the normal observer

will be unable to perceive a figure. In these plates, there are at least three hues that vary in

saturation. The more saturated hues are usually the background colours and the desaturated

colours are usually the figure. The variation in hue throughout the plate masks the difference

in saturation for the colour normal and so the plate design is similar to camouflage. However,

because the background hues are on the same line of confusion for the colour defective and the figure hues are on a different line the color defective sees a desaturated figure within a background of more saturated colour.

Some of PIC tests do not include hidden plates because some normal subjects can read them and the ability to read them depends on the subject's age. For example, 40% of colour-normal subjects between 20 to 30 years can see some hidden figures on the Ishihara test while very few older adults or young children report the hidden figure correctly (Lakowski, 1969b).

1.6 Review of colour vision tests used in this study

Measuring the colour discrimination of people who carry out tasks where colour is used to convey information and accurate colour judgments are essential for safe and efficient performance of the task is important in order to ensure that they can carry out the tasks (Lakowski, 1969a). Today, there is a large number of colour vision tests available 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 is an example of the latter class of tests (Shin et al., 2007).

1.6.1 The efficiency of Ishihara and HRR tests for detecting colour vision deficiency

The Ishihara pseudoisochromatic plates are the most widely used colour vision test to screen for red-green deficiencies. The complete version contains 38 plates: 25 plates contain numerals of one colour embedded in background of a different colour and 13 are designed so that the patient traces a path (Birch, 1993). The numeric plates are divided into

demonstration (plate no.1), transformation (2-9), vanishing (10-17), hidden digit (18-21), and classification (22-25) (Birch, 1993). According to Birch, the mean number of errors for all transformation and vanishing plates is around the maximum of 16 (Birch, 1997). According to her study, the sensitivity and specificity for a failing score of 4 or more errors are 98.7% and 94.1% respectively (Birch, 1997). Birch found that it was difficult to use hidden digit plates for the reasons discussed previously. The sensitivity of these plates was less than 50%. For the classification plates, she reported that 83.2% of protans and 94.1% of deutans were classified correctly based on either missing one figure or identifying one figure as more distinct than the other (Birch, 1997). The most common reason for the misclassification was the inability to classify a subject because they missed both figures on the plates. This means that the classification plates in the 38 plate Ishihara test are reasonably precise in their ability to classify the red-green defects as either protan or deutan.

The HRR pseudoisochromatic test was developed by Hardy, Rand, and Ritter and reprinted in the 4th edition by Richmond Products. It is designed to detect tritan as well red-green colour vision defects and grade their severity. It contains 24 plates that present either one or two coloured symbols (cross, circle, and triangle) within a gray background. From these 24 plates, there are 6 plates for screening purposes (2 plates for blue-yellow defect and 4 for red-green defect) and 14 plates designed to identify the type and severity of the deficiency (10 plates for deutans and protans and 4 plates for tritans). The grading levels are very mild, mild, medium or strong.

Cole et al (2006) reported that the average number of errors made by red-green defects on the HRR screening plates was 4.97(out of six symbols) and no errors on the tritan plates. Relative to the anomaloscope, they found the sensitivity and specificity 1 and 0.96 respectively when the failure criterion was two or more errors on the red-green screening figures. Eighty-six percent of the subjects were classified correctly as protan or deutan. This classification rate was similar to the percentages reported by Birch for the Ishihara. Cole et al found that 31% of the colour vision defectives were graded as mild deficiency, 43% of the subjects as medium, and 26% as severe (Cole et al., 2006). To determine the validity of the severity grading, the HRR was compared with the Farnsworth Munsell (D15) and anomaloscope tests.

The D15 test can grade colour defective subjects into two groups 'pass' and 'fail' indicating that the defect is either mild or moderate-to-severe. Cole et al found that, with one exception, subjects who were classified as mild by HRR passed the D15 test. For those who were classified as medium by HRR, 40% of them failed the D15 test, and 85%, who were classified as severe by the HRR failed the D15.

 In comparison with the anomaloscope, Cole et al found that the average anomaloscope range increased with HRR severity grading. Nevertheless, 5% of the subjects with small range (less than 10 units) on the anomaloscope, which indicates good colour discrimination, were classified as medium or strong on the HRR and 37% of the dichromats were classified as medium; however, none of the dichromats were classified as mild(Cole et al., 2006). This shows that HRR is good for separating severe from mild, but not as effective in separating medium from severe.

19

1.6.2 The Farnsworth Munsell D15 arrangement colour vision test

The Farnsworth Munsell D15 (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). The test consists of a box of 1 fixed reference cap and 15 movable caps of different colours. The subject must arrange these caps according to hue starting from the reference colour on the box.

Birch (2008), conducted a study of 710 men with red-green colour vision deficiency. She used the Nagel anomaloscope to identify the type and severity of the defect (Birch, 2008). According to the study, 53% of the subjects passed the test if one major crossing is allowed. None of the subjects made a tritan error. Four protanopes and two deuteranopes (3% of the dichromats) passed the test using this criterion. Forty percent of the protans and 58% of the deutans passed with the same criterion. The number of dichromats who passed the test increased when more crossings were allowed for a pass (Birch, 2008). In order to avoid passing dichromats, she recommended that:

- Do not allow any major crossings for a pass.
- Retest the subject with 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 errors in the arrangement.

Hovis et al (2004) determined the repeatability of the D 15 using 116 red-green colour defective subjects. They analyzed the repeatability of the number of crossings and three parameters of the Vingrys and King Smith Colour Differences Vectors analyses

(Vingrys & King-Smith, 1988). These parameters were Confusion index(C-index), Specificity index (S-index), and angle size. The C-index indicates the severity of the defect and is correlated with the number of crossings and total error score. The S-index provides measurement of how well the crossings parallel with each other. The angle gives measurement of the type of the defect with protan angles larger than zero and deutan angles smaller than zero (Hovis et al., 2004).

They found that if the failure criterion was 2, or more major crossings, the repeatability of D15 was quite good with kappa (κ) coefficients of 0.84 (Hovis et al., 2004). This corresponded to a C-index greater than or equal to 1.7. Their repeatability value was less than the value calculated from Farnsworth data (96%). The reason for the difference between the two values was that there were a large number of colour-normals in Farnsworth's study, which would improve the repeatability of the test because they did not make any errors at either session.

Their results showed that there was a perfect agreement with the anomaloscope in terms of classifying the type of defect (protan and deutan in both sessions) by visual inspection. They refined the angle value that divides protans from deutans to values greater than -3.0 were classified as protan and values smaller than, or equal to -3.0, were classified as deutan (Hovis et al., 2004). The repeatability was marginally better based on the angle relative to visual inspection with κ values of 0.83 versus 0.77.

1.6.3 Computerized colour vision

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 diagnosis of the severity of the colour vision defect. One of these computer-based tests is the Cone Contrast Sensitivity test (CCST) (Rabin, 2004).

1.6.3.1 Cone contrast sensitivity test

The ability to distinguish a variety of colours is fundamentally dependent on three retinal cone photopigments. Rabin (2004) described a new approach to discriminate the colour vision deficiency based on the photopigments of normal colour vision. He measured the cone contrast sensitivity for letter recognition. The letters were fixed in size and dominant wavelength, but chromatic contrast was varied until the letters could no longer be resolved. The chromatic contrast was varied such that only one of the three cones (S, M, or L) would be able to resolve the letter within the gray surround. This technique allows for a quantitative measure of normal colour vision and a diagnosis as to the type and severity of a colour vision deficiency.

The cone contrast values were determined by the monitor luminance and CIE (x,y) chromaticity coordinates of the three phosphors and then transformed into cones excitation based on psychophysically derived cone spectral sensitivities (Rabin, 1996) . Cone contrast was measured from the amount of cone excitation produced by the coloured letter relative to the gray background (Rabin, 2004). Ten rows of letters were designed for each cone type(S, M, and L). Each row had a contrast change in contrast of 0.1 log unit (Rabin, 2004).

Rabin determined cone contrast sensitivity from 30 colour normal and 28 individuals with a congenital red-green colour vision deficiency. He found that   protan subjects had a clear reduction in L cone contrast sensitivity with normal S and M cone contrast sensitivity. This indicates a direct diagnosis of protan deficiency (Rabin, 2004). Deutan subjects showed

22

decreased in M cone contrast sensitivity with normal S and L cone contrast sensitivity. The

results from cone contrast sensitivity in terms of normal vs abnormal and protan vs deutan

were in perfect agreement with the anomaloscope findings. He found that the severity of

congenital colour vision deficiency based on the cone contrast sensitivity was correlated with

other colour vision tests such as the range of acceptable matches on the anomaloscope (r=0.7,

P<0.001) and the total error score on the FM 100(r=0.65, p<0.001) (Rabin, 2004).

Chapter 2
Purpose

The main objective of this study is to determine the validity and repeatability for a new colour vision test "Colour Vision Reaction Time (CVRT)" for colour vision screening. The study will also determine the repeatability of a selection of clinical colour vision tests, which are currently in use. The clinical colour vision tests are:

➢ **Pseudoisochromatic tests:**

- The 4ᵗʰ edition of the Hardy Rand Rittler test (HRR)
- The 38 plate edition of the Ishihara test (Ishihara)
- Pseudoisochromatic plates Ishihara Compatible (PIPIC).

➢ **Computerized colour vision tests:**

- Prototype ColorDx (pColorDx) Computerized Colour Vision Test (TVC Dayton Military Research Version)
- The Cone Contrast Sensitivity Test (CCST)

➢ **Arrangement colour vision test**

- The Farnsworth Munsell D15

Chapter 3
Subjects

The study recruited 75 colour normal subjects and 47 subjects with a red-green colour vision defect through posters, social media and newsletter advertisements. Colour vision was classified according to the Nagel Anomaloscope. The colour normal participants were 60% females and 40% males whereas the colour abnormal group were predominantly male (94% males and 6% females) because it is X linked-recessive- trait. The subjects ranged in age between 16 years and 71 years and had no known vision problems other than a colour vision problem or a corrected refractive error. The 16 to 71 age range was primarily related to the availability of colour-defectives who were willing to participate in the study. 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 visual acuity of at least 6/6 binocularly with or without spectacles or contact lenses.

This study was received ethics clearance through the office of Research Ethics, at the University of Waterloo (ORE 19211).

Chapter 4
Material and Methods

All tests in this study were administered by the writer. The testing sequences began with a questionnaire (Appendix A) to determine whether the person met the inclusion criteria. For those who had colour vision defect, an additional set of questions was asked regarding when they became aware of the problem and how it affected their lives. Next, binocular distance visual acuity was measured with a Bailey-Lovie chart. The colour vision tests were administered to those individuals who met the inclusion criteria.

The first colour vision test was the anomaloscope. Figure 8 shows the Nagel Anomaloscope. The procedure of this device is based on colour matching. Figure 9 outlines the anomaloscope procedure.

Figure 8. Nagel anomaloscope used in this study
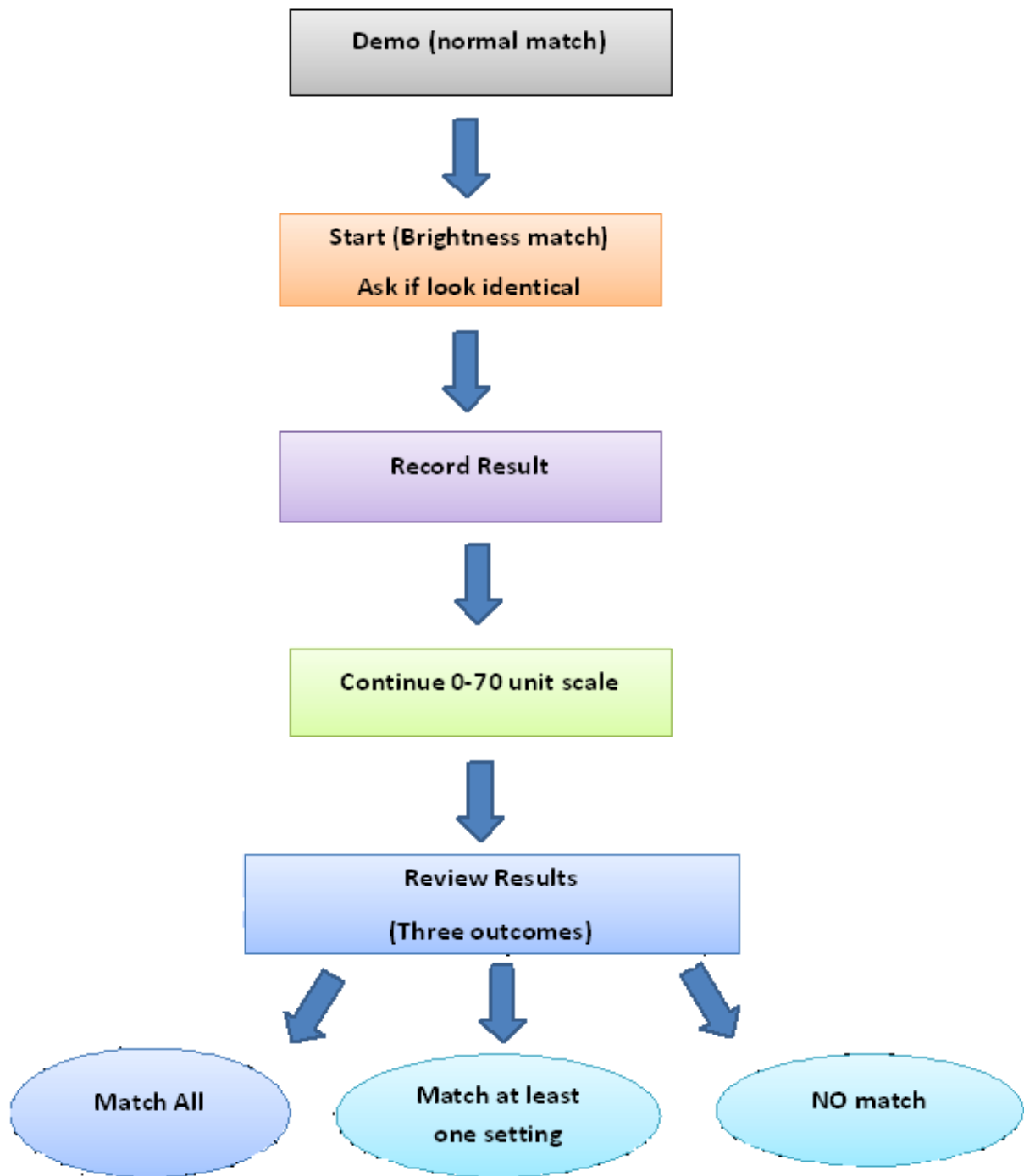
Figure 9. Schematic of the anomaloscope procedure.

Subjects view a circle with different lights presented in the top and bottom halves. The light in the bottom half of the circle is the yellowish monochromatic (i.e. 589nm) light. The top half of the circle contains variable mixture of green (545nm) and red (670 nm) lights. In principle, the subject's task is to adjust the relative amounts of the red and green lights and the brightness of the yellow field so that the two halves of the circle look identical. This was not done in practice because the colour defectives matches tended to be highly variable. The actual procedure was the following:

- The participant was presented with a prepared normal colour match. (For normals 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 in the top field from 0 to 70 unit scale (0 means pure green and 70 means pure red at the top field) in ten unit steps. Their task was to match the brightness of the bottom test 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 matched all scales (full range of mixture was accepted). Subjects were diagnosed as a dichromats (either of protanope or deuteranope based on their brightness match to the red-only light).

   o In this situation, participants were asked to adjust the brightness for the test light to match the brightness with the top field at 70 unit scale (pure red).

- Protanope adjust the test light brightness value to a low value (less than 5) at the red end (70 unit scale) of the mixture field.

- Deuteranope made a small change to the test light brightness value close to the normal's setting.

➢ Second, the subject matched at least one setting. The subjects were diagnosed as normal or anomalous trichromat by determining the exact range by bracketing around the match value.

- Normals accepted the match only at 40. Anomalous trichromats accepted at least one match either below or above 40.

- The range of acceptable matches was determined by presenting the top field in steps with one unit scale below the match point for normals or below the lower setting point of the acceptance for the anomalous trichromat until the subject could notice a difference in colour between the top and bottom field.

- The participant's task here was to adjust the brightness for the bottom field to match the brightness of the top field.

- After the lower limit of the range was determined, the upper limit was measured by presenting the top field in steps with one unit scale above the midpoint for normals or above the upper setting point of the acceptance for the anomalous trichromat until the subject can notice a difference in colour between the top and the bottom field.

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

- If the acceptance match was in the green area (0 to 35 unit scale) the subject was diagnosed as deuteranomalous and the severity determined by the width of the acceptable match range.

- If the acceptance match was in the red area (45 to 70 unit scale) the subject was diagnosed as protanomalous and the severity was determined on the width of the acceptance match range.

➢ Third, if the subject could not make a match, then the subject task was to adjust the red-green values to make a match.

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

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

  o The acceptance range was measured as prescribed previously.

The next series of tests were the pseudoisochromatic plate tests. The common feature for all of these tests was that there was a pattern (numbers or shapes) of coloured dots embedded within a background of differently coloured dots. The three pseudoisochromatic plates tests used in this study were the PIPIC, the 38-plate edition of the Ishihara and the 4th edition of the HRR. Figure 10 shows the PIPIC test, Figure 11 shows the Ishihara test, and Figure 12 shows the HRR test. All of these tests were illuminated with Illuminant C at 1400 lux (±5%) (Minolta T-1 Illuminance Meter, Ramsey, NJ)
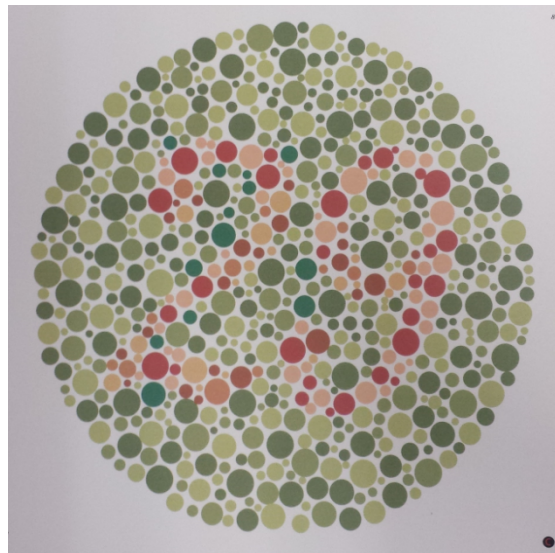
Figure 10. The PIPC test booklet and an example of vanishing plate on the right.
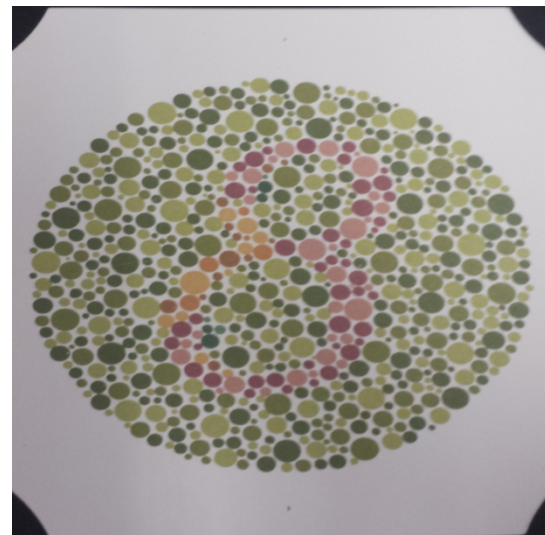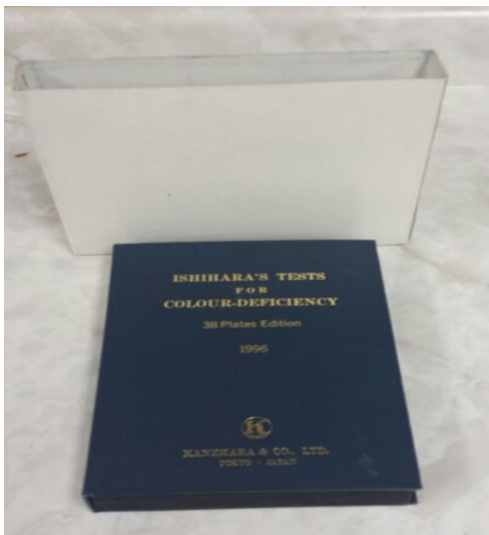
Figure 11. Ishihara test booklet and a transformation plate on the right.

Figure 12. HRR test booklet and example of a vanishing diagnostic plate on the right.

Subjects viewed these tests from approximately 60 cm. The presentation time for each plate was not well controlled, but was approximately five seconds. Each test started with the demonstration plate to make sure the subject understood the test.

The PIPC tests consist of 14 red-green screening plates, one diagnostic plate, and two plates to screen for blue-yellow defects. More than two errors on the red-green screening plates or any error on the blue-yellow plates was considered as a fail. In this test, there is only one diagnostic plate with number 35 on it to determine the type of defect (protan/deutan). Protans would see only a 5 and deutan would report only the 3.

The Ishihara test consists of 20 red-green screening plates and 4 diagnostic plates. If the subject read correctly 17, or more, of screening plates, the colour vision was classified as normal (Birch, 1997). The diagnostic plates were presented to only those who failed the screening series. Deutans read the pink number and missed the purple number. In contrast, subjects with protan defect read the purple number but missed the pink figure.

33

The HRR required the participant to identify geometric shapes (X, Δ, and O) embedded in a background of gray dots. The test started with four demonstration plates followed by two blue-yellow screening plates. Plates 7-10 screened for red-green defects. For the diagnostic purposes, plates 11-20 were presented to determine the type and severity of red-green defect followed by plates 21-24 to determine the blue-yellow defect. Subjects who made no more than one error on the red-green screening plates and none on the red-green diagnostic plates were considered as normal (Cole et al., 2006). In the diagnostic plates, a subject was diagnosed as protan if the total numbers of error in the protan column was fewer than the number of errors in the deutan column, whereas subjects were diagnosed with deutan defect if there were fewer errors in the deutan column. The score sheet in Figure 13 illustrates how the severity was determined.

Figure 13. Recording sheet for HRR test.

After completing the plate tests, three computerized colour vision tests were administered. Two of them (CVRT and pColorDx) were presented on a LG monitor (Model: W2442PAT) using a PC computer with a Windows 7 Professional Operation system. The monitor was calibrated using Spyder program (4PRO 4.4.5 Version) to a white reference of 6500° K correlated colour temperature.

The first computer based test was Colour Vision Reaction Time (CVRT). This was designed to screen for red-green and blue-yellow defect. The principle of the CVRT was to measure the response times for targets of different chromaticities. The CVRT stimulus

consists of a small square-shaped target presented within a green background. Both the target and the background are composed of smaller squares, which vary randomly in size and luminance. This creates both luminance and contour noise so that the subject has to identify the location of the target square based on hue differences. The test was programed using MATLAB (MATLAB R2012b 32-bit). There are three different test hues. White was used for a control, yellow was used to screen for red-green defects and blue-green was used to screen for a blue-yellow defect. Figure 14 shows a screen shot of the CVRT with the white control target presented. Figure 15 shows the CIE diagram for the chromaticity coordinate used in this test. As can be seen, the yellow target and the green background fall on the same red-green line of confusion (orange line) and the blue-green target falls on a tritan line of confusion with the green background (blue line).
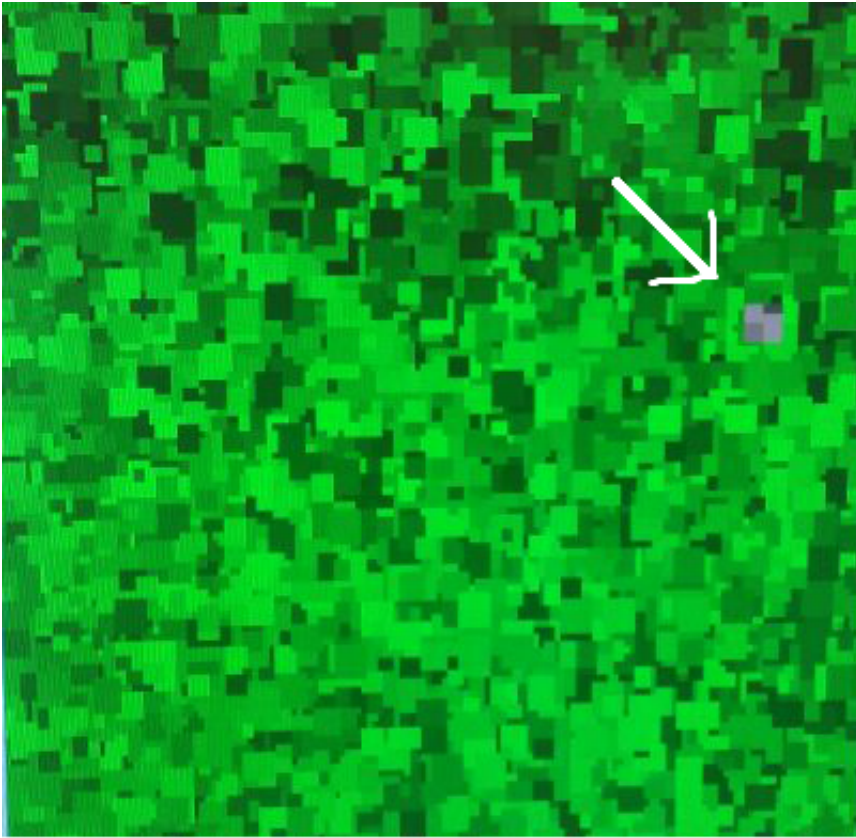
Figure 14. Screen shot of the CVRT for a gray target.
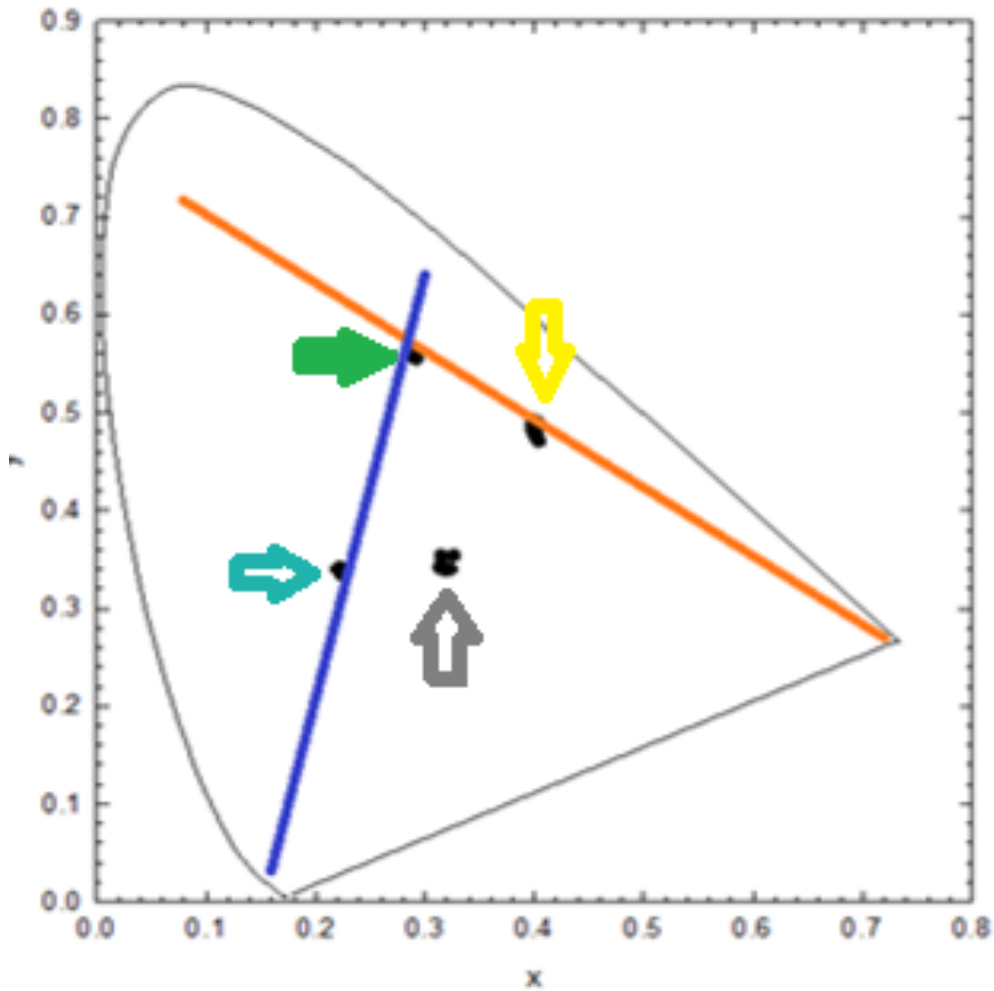
Figure 15. CIE diagram for CVRT test. The yellow, gray, and blue-green arrows indicate the chromaticity coordinates for yellow, gray, and blue-green targets. The sold green arrow indicates the chromaticity coordinates for the green background.

The participant's task was to identify the position of the square target by moving the computer mouse onto the target and clicking one of the mouse buttons. Both colour-normals and colour-defectives were able to see the white target easily, whereas the colour defective would either not be able to locate either the yellow or blue target because it was indistinguishable from the background or would take longer to respond correctly.

The test starts with eight practice trials using the white target. Some of the practice trials have blank trials (where no target is presented) to remind subjects that some of the screens may have no target presented. This done in order to decrease the expectation that target will be presented on every presentation and reduce random guessing. More practice trials are allowed if requested by the subject. Experimental trials consist of displaying a green background with a target of one of two test hues or the white control hue. A gray background with squares of randomly varying luminance is presented between trials to maintain a neutral adaptation state. A warning signal in the form of a three short "beeps" sounds is presented to alert the participant that the next trial will be beginning soon. After the sound is finished, there is a 1-3 second delay before the target and green background images are presented. The time taken to identify the location of the target with the mouse pointer is recorded if the location is identified correctly within a 5 sec presentation period. If the target square is not identified correctly within this time, then a miss is recorded and 5 sec is arbitrarily recorded as the reaction time. The experiment presented a total of up to 30 trials (15 control targets and 15 targets of a given hue); however if the subject misses 8 trials of the test hue, then only the white target is presented for the remaining 15 trials. The yellow target hue was always presented before the blue-green target. The order of the control and target presentation was determined by a pseudo-random sequence. At the end of the session, the participant's reactions times were presented all of the trials, as well as the calculated mean reaction time and standard deviation for each hue and the number of misses.

After the CVRT test, the Farnsworth Munsell D15 was administered. In this test, the participant arranges different colour caps starting from the first fixed colour on the box as

shown in Figure 16. Participants were asked to place the coloured cap that was most similar to the last one placed in the box. The test was administered three times without a feedback.

Figure 16. D15 arrangement test. Red arrow indicates the reference cap.

The pColorDx was the second computerized colour vision test administered. This test presents numbers using the pseudoisochromatic design. Each plate was presented for 2 sec within a white background. After the number disappears, a list of nine black numbers appeared on the screen and the subject selected which of the nine numbers was on the plate. The test presented up to 25 screening images for red-green defects followed by 12 screening plates for the tritan defect. However, the red-green screening test ended once a total of 5 errors were made and the program switches to the diagnostic series starting with the blue-yellow series. After the screening plates and blue-yellow test plates, 64 diagnostic plates were administered for individuals who failed the red-green screening series. Half of the plates had figure colours that were along the protan line of confusion and the other half were

along the deutan line of confusion. The protan series was presented before the deutan series. The saturation varied in each series so that the plates estimated the severity of the defect by estimating the discrimination threshold from the gray for each red-green defect. The subject was classified as protan if the total numbers of errors on the protan series were more than the deutan, and as a deutan if the total numbers of errors on the deutan series were more than the protan series. Table 2 shows the classification of the severity for red-green defect on the pColorDx test.

Table 2. The classification of severity for red-green defect (number of errors on the diagnostic plate on pColorDx test.

MILD	More than 5 & less than 17 errors
MODERATE	More than 17 & less than 28 errors
SEVERE	More than 28 errors

The last test administered in the sequence was the Cone Contrast Sensitivity Test (CCST). This test presents coloured letters within a gray background. Three different coloured letters were generated for each cone type (L, M, and S cone). In this study, we created our own version by displaying the letter on a CRT screen monitor (SONY model GDM-F520) using PowerPoint slide presentation driven by laptop as shown in Figure 17.
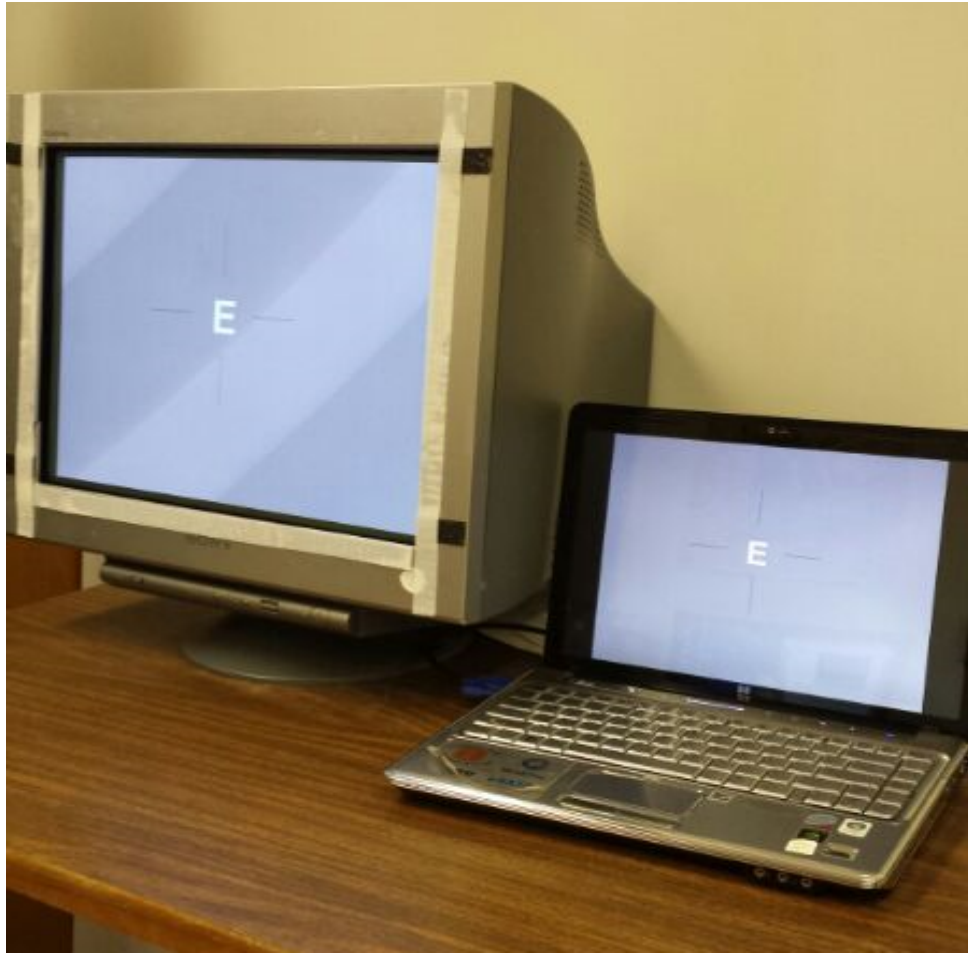
Figure 17. Cone Contrast test. The monitor on the left was the test display. The laptop on the right controlled the test.

The different cone contrasts were established by setting the RGB values of the letter so that CIE x, y and Y values produced on the CRT monitor matched cone contrast supplied by Dr Jeff Rabin. Figure 18 shows the CIE values for CCT used in this study and Table 3 shows the actual and desired cone contrast values supplied by Dr Rabin. The chromaticity coordinates for all coloured letters and luminance were measured by luminance photometer (Colorcal colorimeter ''Konica-Minolta, Ramsey. NJ").
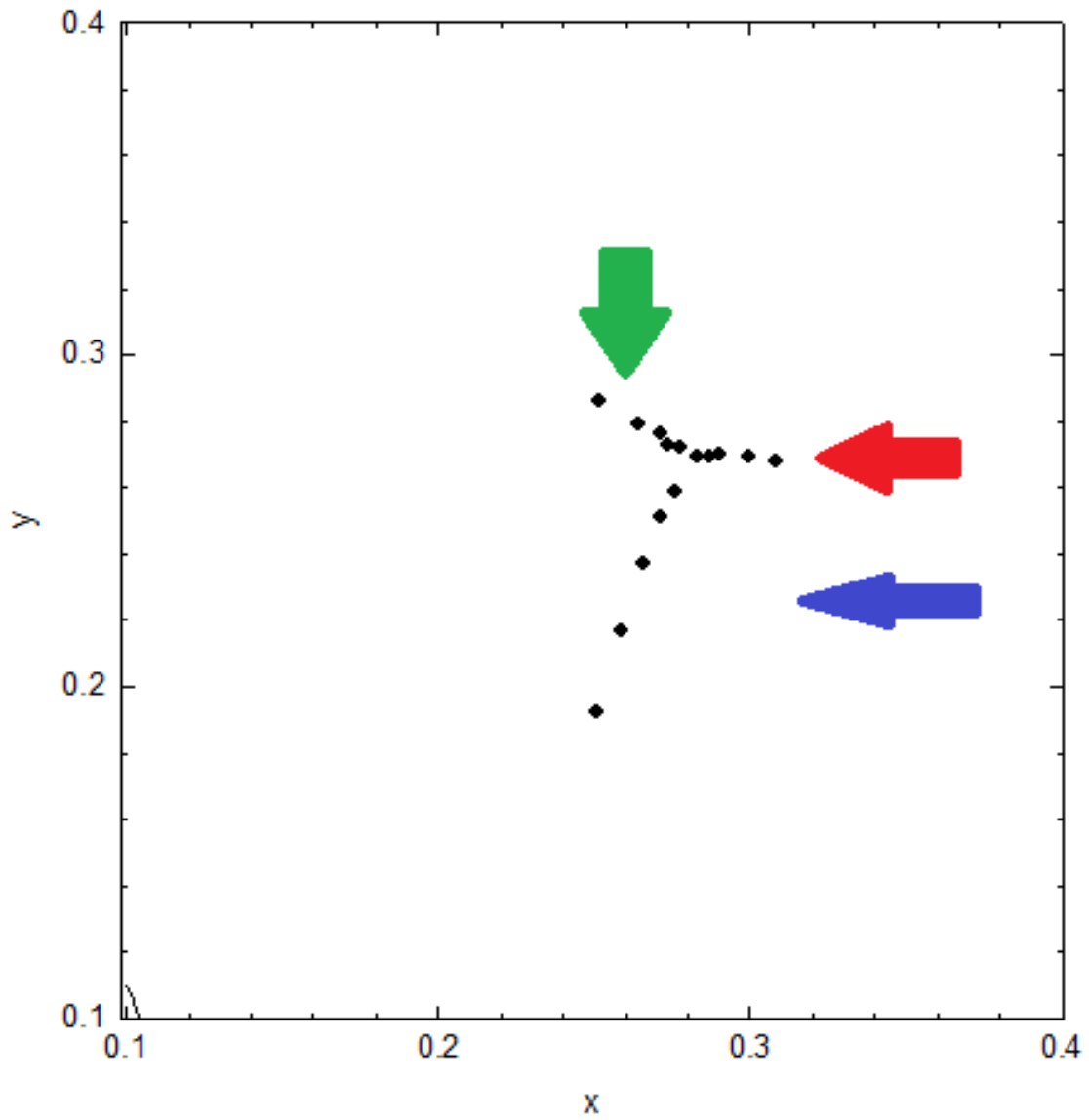
Figure 18. CIE values for CCT. Green, red, and blue arrows indicate the chromaticity coordinates used in this study for M cone, L cone, and S cone contrast respectively.

.

Table 3. The desire cone contrast values used in Jeff Rabin study and the actual cone contrast values used in this study.

	Desired CC	**Actual CC**
L Cone Contrast	9.12	9.32
	5.25	5.27
	3.02	2.94
	1.74	1.73
	1.00	1.06
M Cone Contrast	9.12	9.37
	5.25	5.66
	3.02	3.14
	1.74	1.64
	1.00	1.01
S Cone Contrast	72.44	73.8
	41.69	44.9
	23.99	25.4
	13.8	13.46
	7.94	7.48

There were five cone contrast values for each series of letters and five letters presented at each contrast. One letter at a time was presented for two seconds and the participant's task was to name the letter. The test presented the letters in decreasing contrast. All 5 letters were presented before going to the next contrast level. The test was scored

similar to the Peli Robson chart where each letter is assigned a value of 0.05 log contrast unit. The individual cone contrast sensitivities are calculated by multiplying 0.05 by the number of correct letter. The size of the letter was 70 point font and when viewed from the test distance of 1.3 meters was equivalent to a 6/90(20/320) sized letter.

Following the recommendation of working group 41 we will use the kappa (κ) coefficient of agreements (the correct equation was used) to determine the validity and repeatability of the tests (Working Group 41, 1981). We used the 95% confidence interval of κ coefficient to determine statistical significance.

Chapter 5
Results

5.1 Questionnaire

All subjects, other than wearing spectacles or contact lenses, having dry eye, and/or having a colour vision problem, were not aware of any vision problems or problems with their eyes nor were they being treated for any eye diseases. Interestingly, only 73% of the colour–normal subjects were sure about their normal colour vision.

Twenty eight percent of the colour defective subjects were unsure as to whether they had a colour vision defect. The remaining 72% reported that they had a defect. This percentage was slightly lower than the 82% value reported by Steward and Cole (Steward & Cole, 1988). Of the subjects who were certain about their defect, 6% were aware prior to any schooling and the others became aware of the defect between ages of 6 to 25 years. Although they were sure about the defect, 67% did not report any independent confirmation. The others were informed by an optometrist (6%), a parent (15%), a teacher (3%), an army physician (3%), and a university course (6%). Eleven percent of the colour defective subjects reported that their defect influenced their career choice.

Thirty four colour defective subjects (72% of the colour defectives) were asked to describe the colour of the little man figure in the pedestrian traffic light indicating that it is safe to cross the street. Seventy one percent of them reported that figure was green. The probable reason for the misnaming is that the white colour of the figure is on the same line of

confusion with the green colour in the traffic light, and within the context of traffic, both colours would be identified as green.

5.2 Pseudoisochromatic plates

5.2.1 HRR Test

5.2.1.1 Screening plates

Most of the colour defective subjects made errors on the red-green screening plates. The two exceptions were one deuteranomalous subject at the first visit and two deuteranomalous individuals (one was the previously mentioned subject) at the second visit. The average number of errors out of 6 possible responses on the red-green screening plates was 5.05 at the first visit and 4.93 at the second visit. There were 2 deuteranomalous subjects at the first visit and one different deuteranomalous subject at the second visit who made only one error on the tritan screening plates. One subject (from the three) missed the green yellow circle on the first plate and the other two missed the purple triangle on the second plate. None of these subjects missed any of the tritan diagnostic plates.

Using Cole, et al's pass-fail criteria (no more than one error on the red-green figures for as pass), only one colour defective subject passed the screening plates at the first visit and two passed at the second visit. None of the colour-normals failed the red-green screening plates. Table 4 shows the comparison between the HRR screening plates and the anomaloscope at the first visit. There was a good agreement with anomaloscope with a kappa coefficient agreement (κ) equal to of 0.98 (standard error± 0.017) at the first visit. The sensitivity and the specificity were 99% and 100% respectively.

Table 4. Comparison between the HRR red-green screening plates and the anomaloscope

		Anomaloscope	
		Pass	**Fail**
HRR	**Pass**	75	1
	Fail	0	46
		Specificity 100%	**Sensitivity 99%**
Kappa Coefficient		**0.98 standard error ± 0.017**	

Table 5 shows the repeatability of the HRR red-green screening plates, which was very good with κ coefficient of agreement equal to 0.98 (standard error of ± 0.017). The single discrepancy between the two sessions was a result of one deuteranomalous subject who failed the screening plates at the first visit but passed at the second visit.

Table 5. Repeatability of the HRR red-green screening plates.

		HRR 1st session	
		PASS	**FAIL**
HRR 2nd session	**PASS**	76	1
	FAIL	0	45

5.2.1.2 Diagnostic plates

Of the individuals who failed the red-green screening plates, the HRR correctly classified 100% of the red-green colour defective as either protan or deutan at both visits. Table 6

shows the HRR classification of the subjects as protan, deutan, or unclassified related to the anomaloscope. The agreement with anomaloscope (in terms of classifying the defect as either deutan or protan) and repeatability of the classification was perfect with κ=1.0 (standard error of ± 0).

Table 6. HRR classification relative to the anomaloscope.

		ANOMALOSCOPE	
		PROTAN	DEUTAN
HRR	**PROTAN**	14	0
	DEUTAN	0	32
	UNCLASSIFIED	0	0

5.2.1.3 Grading the severity

Figure 19 shows the relationship between the HRR severity classification and the range of acceptable matches for the anomaloscope. The mean values from Cole, et al's (2006) study are included for comparison. The figure shows an increase in the anomaloscope range with severity rating on the HRR. The Spearman rank correlation coefficient of 0.6 was significant (p<0.0001), but only moderate.

One of the reasons for the less than perfect correlation was that 27% of the dichromats (14% deuteranopes and 62% of the protanopes) were classified as moderate instead of severe. There were also two deuteranomalous subjects who would be considered to have a mild colour vision defect based on a relatively small matching range (i.e. <20 units) who were classified as moderate or severe on the HRR.
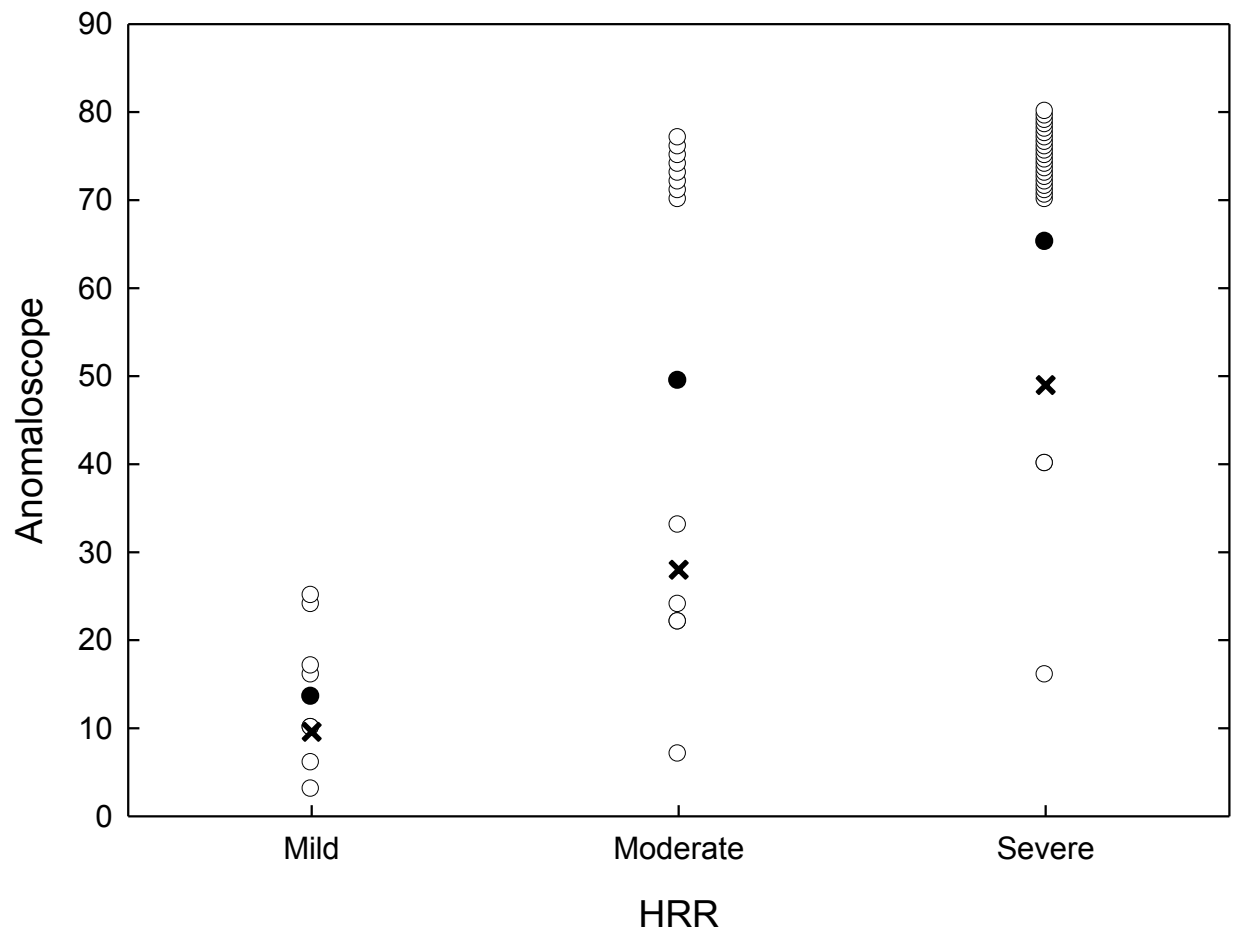
Figure 19. Relationship between the HRR severity classification and the anomaloscope range of acceptable matches for the colour defective subjects (individual dichromats ranges have been offset to show the number of individuals at each severity grade). Solid circles indicate the average matching range in the anomaloscope used in this study. The black x's indicate the mean range results from Cole, et al's study.

Table 7 shows the repeatability of the HRR severity classification. There was a good agreement between the first and second visit with κ coefficient of 0.73 (standard error of ± 0.092). There were 7 subjects who had a different classification of their severity between

visits. Two of the subjects (both deuteranopes) had a more severe classification at the second session and 4 individuals (three dichromats and one anomalous trichromat) improved from a severe classification to moderate at the second session.

Table 7. Number of subject with colour vision deficiency classified as mild, moderate and severe by HRR in the first and second visit.

		HRR 1st Session		
		MILD	MODERATE	SEVERE
HRR 2nd Session	**MILD**	6	0	0
	MODERATE	1	12	4
	SEVERE	0	2	20

5.2.2 Ishihara Test

5.2.2.1 Screening plates

The average number of errors made by the colour defective subjects on the screening plates was 17.6 at both visits. Using more than 3 errors on transformation and vanishing plates as a failure (Birch, 1993), 96% of the colour defective subjects failed the test at the first visit and 98% the colour defective subjects failed the test at the second visit. Two deuteranomalous passed the screening plates at first visit and one at the second visit. All the colour-normals passed the screening plates at both visits. Table 8 shows the comparison between the Ishihara screening plates and the anomaloscope at the first visit. There was a good agreement with

anomaloscope with κ coefficient of 0.96 (standard error± 0.02) at the first visit. The specificity and the sensitivity were 100% and 96% respectively.

Table 8. Comparison between the Ishihara screening plates and the anomaloscope.

		Anomaloscope	
		Pass	Fail
Ishihara	Pass	75	2
	Fail	0	45
		Specificity 100%	Sensitivity 96%
Kappa coefficient		0.96 (standard error± 0.02)	

Table 9 shows the repeatability of the Ishihara screening test. The repeatability between the first and second visit was very good with the κ coefficient of 0.98 (standard error of ± 0.017). The one discrepancy was the result of one deuteranomalous subject who failed the screening plates at the first visit but passed it at the second visit. The other deuteranomalous passed the Ishihara test at both sessions.

Table 9. Repeatability of the Ishihara screening plates.

		ISHIHARA 1st Session	
		PASS	FAIL
ISHIHARA 2nd Session	PASS	77	1
	FAIL	0	44

5.2.2.2 Diagnostic plates

The diagnostic plates classified correctly 76% of the deutans and 86% of protans at the first

visit. Five deutan subjects and 2 protan subjects were unclassified. This was because they did

not make any errors on the diagnostic plates. There were no misclassifications by the

diagnostic plates. Table 10 shows the Ishihara classification results as a function of the

different types of colour vision defects and Table 11 shows the comparison of the diagnostic

plates with the anomaloscope for the first visit. The agreement with the anomaloscope was

fair with κ coefficient of 0.73 (standard error ± 0.089). The lower agreement between the

diagnostic plates and the anomaloscope was a result of the unclassified anomalous

trichromats.

**Table 10. Percentage of subjects who were classified by the Ishihara test as deutan or protan in
comparison with the Nagel anomaloscope diagnosis.**

	First visit	
	Correctly classified	**unclassified**
Deuteranope	100%	0
Deuteranomalous	55%	45%
Protanope	100%	0
Protanomalous	67%	33%

Table 11. Ishihara classification relative to the anomaloscope diagnosis.

		ANOMALOSCOPE	
		PROTAN	DEUTAN
ISHIHARA	PROTAN	12	0
	DEUTAN	0	27
	UNCLASSIFIED	2	4

Table 12 shows the agreement in the diagnostic plates between the first and second visit. In terms of the repeatability, there was a good agreement between the first and second visit with κ coefficient of 0.84 (standard error ± 0.071). The discrepancies between both visits were because

- 2% of the colour defectives were unclassified at the first visit but classified correctly as protan in the second visit.

- 7% of the colour defectives were classified as deutan at the first visit but were unclassified (see both figures) at the second visit (2 deuteranomalous and 1 deuteranope).

Table 12. Repeatability of the Ishihara diagnostic plates.

		ISHIHARA 1st Session		
		PROTAN	DEUTAN	UNCLASSIFIED
ISHIHARA 2nd Session	PROTAN	12	0	1
	DEUTAN	0	24	0
	UNCLASSIFIED	0	3	5

5.2.3 Pseudoisochromatic plates Ishihara Compatible (PIPIC)

5.2.3.1 Screening plates

The average number of errors made by the colour defective subjects on the 14 screening

plates at both visits was 12. Using the recommended (Waggoner, 2003) failure criterion of

more than 2 errors, 96% of the colour defectives failed the test at the first visit and 92%

failed the test at the second visit. Table 13 shows the comparison between the PIPC screening

plates and the anomaloscope at the first visit. There was a good agreement with

anomaloscope with coefficient κ equal to 0.96 (standard error ± 0.02) at the first visit. The

specificity and the sensitivity were 96% and 100% respectively.

In terms of blue-yellow defect, only one deuteranomalous subject failed the blue-

yellow screening with two errors. This subject was one of the subjects who also made errors

on the HRR blue-yellow screening plates. In terms of the repeatability, all the subjects

(normals and colour defectives) passed the blue-yellow plates at second session without

errors.

Table 13. Comparison between the PIPC screening plates and the anomaloscope.

		Anomaloscope	
		Pass	**Fail**
PIPC	**Pass**	75	2
	Fail	0	45
		Specificity 100%	**Sensitivity 96%**
Kappa coefficient		0.96 standard error ± 0.02	

Table 14 shows the repeatability of the PIPC screening test. There was a good agreement between the first and second visit with κ coefficient of 0.96 (standard error of ± 0.024). The small number of discrepancy was two deuteranomalous subjects who failed the screening plates at the first visit but passed at the second visit.

Table 14. Repeatability of the PIPC screening plates.

		PIPC 1st Session	
		PASS	**FAIL**
PIPC 2nd Session	**PASS**	77	2
	FAIL	0	43

5.2.3.2 Diagnostic plates

Sixty-seven percent of the deutans and 86% of the protan were classified correctly by the diagnostic plates at the first visit. Table 15 shows the PIPC classification percentages as a function of the colour vision defect. The colour defective subjects who were unclassified either missed both figures (4 deuteranopes and 9 anomalous trichromats) or read correctly the numbers on both plates.

Table 15. Percentage of subjects who were classified in the PIPC test as either deutan or protan as a function of their colour vision defect.

	First visit	
	Correctly classified	**Unclassified**
Deuteranope	78%	22%
Deuteranomalous	28%	72%
Protanope	100%	0
Protanomalous	67%	33%

Table 16 shows how the PIPC classification in relation to the anomaloscope. The agreement with the anomaloscope was fair with κ coefficient of 0.53 (standard error ± 0.088) at the first visit. The lower agreement between the diagnostic plates and the anomaloscope was essentially a result of 28% of the colour defective subjects at the first visit were unclassified by the PIPC test.

Table 16. PIPC classification related to the anomaloscope.

		ANAMOLOSCOPE	
		PROTAN	**DEUTAN**
PIPC	**PROTAN**	12	0
	DEUTAN	0	20
	UNCLASSIFIED	2	11

Table 17 shows the repeatability of the diagnostic plates. There was reasonable agreement between the first and second visit with a κ coefficient of 0.77 (standard error ± 0.083). The discrepancies between both visits were because

- 12% of the colour defectives were unclassified at the first visit but were classified correctly as deutan at the second visit.

- 2% of the colour defectives were classified as deutan at the first visit but were unclassified (saw both figures correctly) at the second visit.

Table 17. Repeatability of the PIPC diagnostic plates.

		PIPC 1st Session		
		PROTAN	**DEUTAN**	**UNCLASSIFIED**
PIPC 2nd Session	**PROTAN**	12	0	0
	DEUTAN	0	19	5
	UNCLASSIFIED	0	1	6

5.2.4 General discussion

The current study confirms that the three pseudoisochromatic tests are effective in screening for red-green colour vision defect. Table 18 summarizes the κ values of agreement for screening red-green colour vision defects with the specificity and sensitivity of each test from the first session. All colour-normal subjects passed the three tests with 100% specificity. Although not statically significant based on the 95% confidence intervals, the HRR had the highest level of agreement with anomaloscope and sensitivity. The Ishihara and PIPC tests screening results were identical, with the same deuteranomalous subjects passing both tests.

Table 18. Level of agreement with anomaloscope and 95% confidence interval, sensitivity and specificity for the three pseudoisochromatic tests evaluated in this study.

	HRR	ISHIHARA	PIPC
Kappa Coefficient	0.98	0.96	0.96
95% CI	0.95-1.00	0.92-1.00	0.92-1.00
Sensitivity	99%	96%	96%
Specificity	100%	100%	100%

Table 18 summarizes the κ coefficient for the three tests in terms of the screening repeatability. All the dichromats failed the red-green screening plates in the three tests at both visits. The κ coefficient values were significantly identical based on the 95% confidence intervals. The validity and repeatability of the Ishihara was similar to the values of earlier studies summarized by Working Group 41 (1981).The HRR may be preferred over the Ishihara and PIPC because sensitivity was marginally higher than the other two tests.

Table 19. Repeatability of the three pseudoisochromatic tests.

	HRR	ISHIHARA	PIPC
Kappa Coefficient	0.98	0.98	0.96
95% CI	0.95 to 1.00	0.95 to 1.00	0.92 to 1.00

Table 20 shows the agreement values for the three tests with anomaloscope in terms of classifying the defect. Agreement of the diagnostic plates with the Nagel anomaloscope in terms of the type of red-green defect varied across tests. The agreement with the Nagel anomaloscope was perfect with the HRR test and acceptable with the Ishihara, but only fair

for the PIPC test. The lower agreement for the Ishihara and PIPC test was likely due to the limited numbers of the diagnostic plates in the PIPC test (only one plate) and in the Ishihara test (four diagnostic plates) combined with our procedure of not asking the subjects which figure was more visible if they reported both figures on the diagnostic plates. On the Ishihara test, our study classified 76% of the deutan and 86% of the protans correctly whereas in Birch (1997) 83% of the protans and 82% of the deutans were correctly classified. She had a better classification in terms of the deutans because her procedures were to ask which figure was more visible if they reported both figures.

However, HRR classification results were slightly better than reported by Cole, et al (2006). There was 100% agreement in our study as to whether the colour defective subjects were classified correctly as protan or deutan, whereas Cole at al reported that only 86% of the colour defective were correctly classified as protan or deutan. This may have been a result of their colour defective sample having a higher percentage of individuals with milder defects. These individuals may have been less likely to miss any diagnostic figures.

Table 20. The Kappa coefficient values for agreement with Nagel anomaloscope in terms of classifying the colour vision defect as protan or deutan.

	HRR	ISHIHARA	PIPC
κ coefficient	1	0.73	0.53
95% CI	1	0.55 to 0.91	0.35 to 0.70

Table 21 summarizes the classification repeatability of the diagnostic plates for the three pseudoisochromatic tests. The agreement between the first and second visit for

classification was perfect for the HRR test and good for the Ishihara test, whereas it was only
reasonable for the PIPC test.

Table 21. Summarized the repeatability of the three tests in terms of the diagnostic plates.

	HRR	**ISHIHARA**	**PIPC**
κ coefficient	1	0.84	0.77
95% CI	1	0.70 to 0.99	0.61 to 0.94

The HRR test was designed to classify the severity of the defect and the 4th edition
was revised in an attempt to provide a better classification of the severity (Bailey et al, 2004).
Our results were similar to Cole, et al's in that there were a reasonable correlation between
the HRR severity and the Nagel anomaloscope matching ranging. The reason that our mean
matching ranges in Figure 1 were higher than Cole, et al's values was probably because they
had a large percentage of colour defective subjects with milder defect.

The κ coefficient for the repeatability of the HRR severity classification was 0.73.
The reason for the lower agreement was the small number of colour defective subjects who
were unclassified at the first session but classified correctly as either protan or deutan at the
second visit.

The HRR and PIPC tests can screen for blue-yellow defects. In the HRR test, all the
subjects (colour normal and colour defective) passed the blue-yellow screening plates at both
sessions if there was only a single error was allowed. However, there were two
deuteranomalous at the first visit and one different deuteranomalous at the second visit who

had a single error on the blue-yellow screening plates, but they did not make any single error on the blue-yellow classification plates.

With the exception of one deuteranomalous subject, all the other subjects (normals and colour-defective) passed the PIPC blue-yellow screening plates (any error considered as fail). This deuteranomalous individual failed to identify both plates at the first visit, but passed it (both plates) at the second visit. This individual also made one error on the HRR blue-yellow screening plates at the first visit, but did not make any blue-yellow errors at the second visit.

We do not believe that these subjects have a blue-yellow defect in addition to the red-green defect based on their visual acuity and brief history. It is possible that these individuals may have a discrimination ellipse around the gray background that has either a longer minor axis than the average deuteranomalous individuals or the long axis of the ellipse is rotated slightly toward the tritan lines of confusion. It is also possible that these mistakes could be an error of expectation in that the subject was expecting not to perceive any figure or symbol and did not examine the plate carefully. The lack of repeatability of the blue-yellow errors for some individuals supports this last hypothesis.

5.3 Computerized colour vision tests

5.3.1 Colour Vision Reaction Time (CVRT) test

5.3.1.1 Red-green screening

None of the colour normal subjects missed the yellow or white target in the red-green screening test. The average response time of the colour normal for white control at the first

session was 1.11 sec (SD ± 0.15) and 1.12 sec (SD ± 0.14) for the yellow. The responses

were a little faster at the second session with 1.08 sec (SD ± 0.11) for the white control and

1.10 sec (SD ± 0.12) for the yellow. The mean response times to locate the white control for

the color-defectives were 1.17 sec (se ± 0.03) at the first session and 1.18sec (se ± 0.03) at

the second session. The mean difference in response times between the white control and

yellow figure for the colour normal was 0.012 sec (95%CI -0.17 to 0.041). Recall that the test

ends when 8 of the yellow targets are missed. All the dichromats subjects and 44% of the

anomalous trichromats failed the red-green screening based on this criterion. The remaining

23% of the colour-defective sample identified the location of the yellow square on more than

50% of the trials and 73% of these individuals identified the yellow target location correctly

on every trial.

In order to improve the sensitivity of the CVRT test, a number of different pass/fail

criteria were examined. All were based on combinations of errors and response times.

Although none of the colour normals missed any yellow targets, we set the miss level for a

failure at 2, or more in case colour normal did have a lapse and missed a target. Using just

this criterion, the sensitivity was 83 % and the specificity was 100%. Next, the response

times for the yellow target were examined. For analysis, colour-defectives who made less

than 8 errors were included to ensure that there was an adequate statistical power. The

individual mean yellow response times were expressed as *z-scores* relative to the white

control. This value was the difference in the subject's mean white control response and the

yellow response times divided by the standard deviation of the white control. Figure 20

shows the Receiver Operating Characteristic (ROC) curve for the CVRT test with different

cut off points. The optimum was 2.22 units. For convenience, this was rounded to 2.2 units.
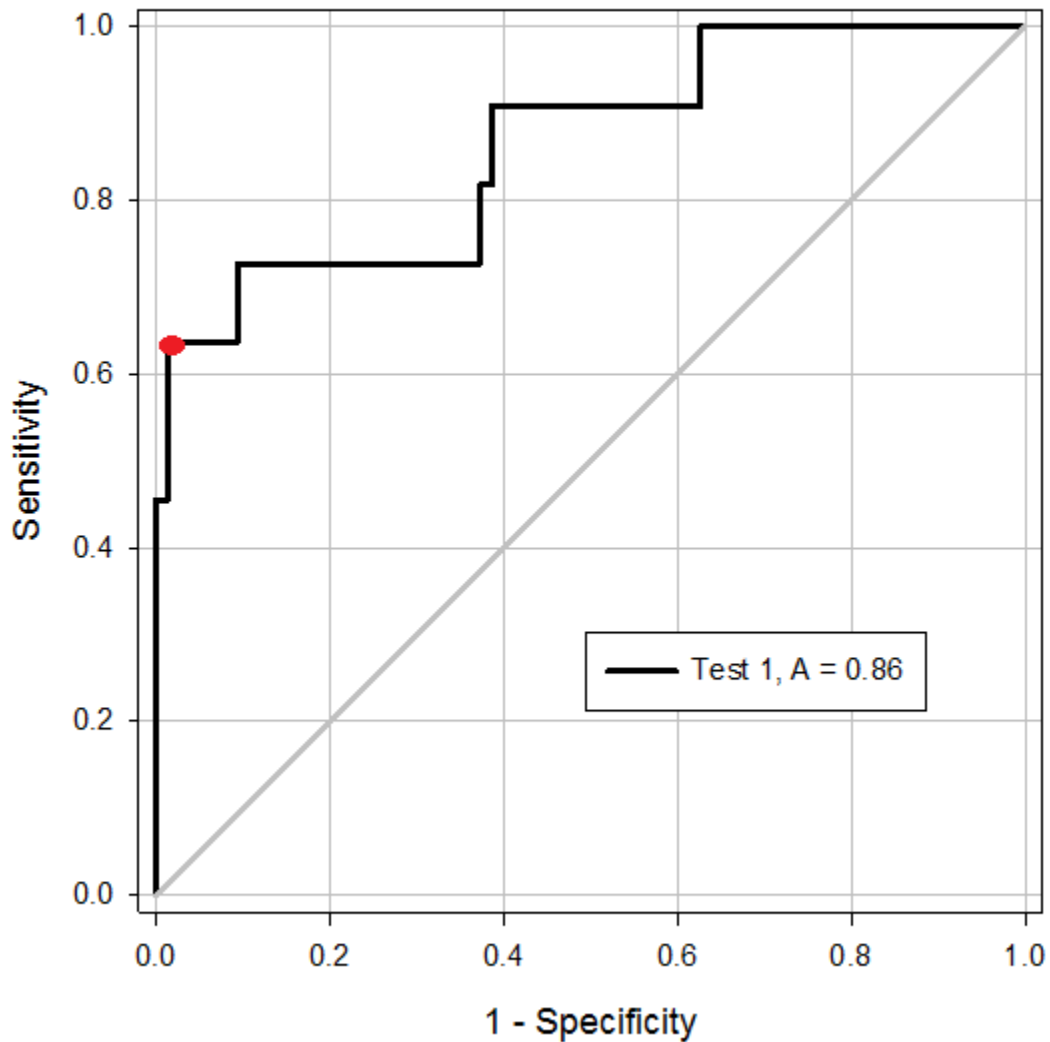
Figure 20. Receiver Operating Characteristic (ROC curve) for the CVRT test. (A) indicates the area under the curve. The red circle indicates the cut-off point (2.2 units) that has the highest sensitivity and specificity.

Table 22 shows the comparison of the CVRT with anomaloscope using two or more misses or a yellow response time score greater than 2.2 units from the white control. There was a good agreement with κ coefficient of 0.91 (standard error= ±0.038). The discrepancies between the two tests were because

- 1.4% of the colour normal failed the CVRT test.

- 8.5% of the colour defective (all deuteranomalous) passed the CVRT test.

Table 22. Comparison between the CVRT test and the anomaloscope.

		Anomaloscope	
		Pass	Fail
CVRT	**Pass**	74	4
	Fail	1	43
Sensitivity		**Specificity 98.5%**	**Sensitivity 91.5%**
Kappa coefficient		0.91 standard error= 0.038	

Table 23 shows the repeatability of the CVRT. There was a good agreement between the CVRT test first and second session with κ coefficient of 0.98 (standard error= 0.017). The one discrepancy between the two sessions was the colour-normal who failed the test at the first session but passed at the second session.

Table 23. Repeatability of the CVRT test.

		CVRT 1	
		PASS	**FAIL**
CVRT 2	**PASS**	78	1
	FAIL	0	43

5.3.1.2 Blue-yellow screening

All the colour-normal subjects and colour defective subjects passed the blue-green test without any misses reported. Their response times to the blue-green figures (except one deutanomalous) were within a z-score of ±1 relative to the white control. The one deuteranomalous had z-score of 2.7 and was the same subject who missed one blue-yellow figure on the HRR test and was discussed previously. All blue-green z-scores were within ±1.0 of the white control at the second visit

5.3.2 Cone contrast sensitivity test

Figure 21 shows the cone contrast results for colour-normal and colour defective subjects. The black box indicates the mean value and the error bars are three standard deviations from the mean (±3 SD form the mean) for colour normal subjects. The ±3 SD was used as a cut-off in previous studies (Rabin, 2004). Solid and white triangles present protan and deutan for all colour defective subjects. There was clear reduction in L cone contrast sensitivity for the protans and M cone contrast sensitivity for the deutans. However, some deutans also showed a reduction in L-cone sensitivity and nearly all the protans showed a reduction in M cone

contrast sensitivity, although reduction in their M-cone sensitivity was not as large as the reduction in their L-cone sensitivity.

In terms of the S cone contrast sensitivity, there were three deutans who had a cone contrast sensitivity value below 0.6 ($\pm$3 SD from the colour normals mean). In comparison with HRR test, all three deutans passed the blue-yellow screening plates without any error. One of the three deutans also failed the pColorDx blue-yellow test at both sessions with 7 errors. However, all three deutans showed improvement in their S-cone contrast sensitivity at the second session so that they were within $\pm$3 SD from the colour normal mean of the second session.
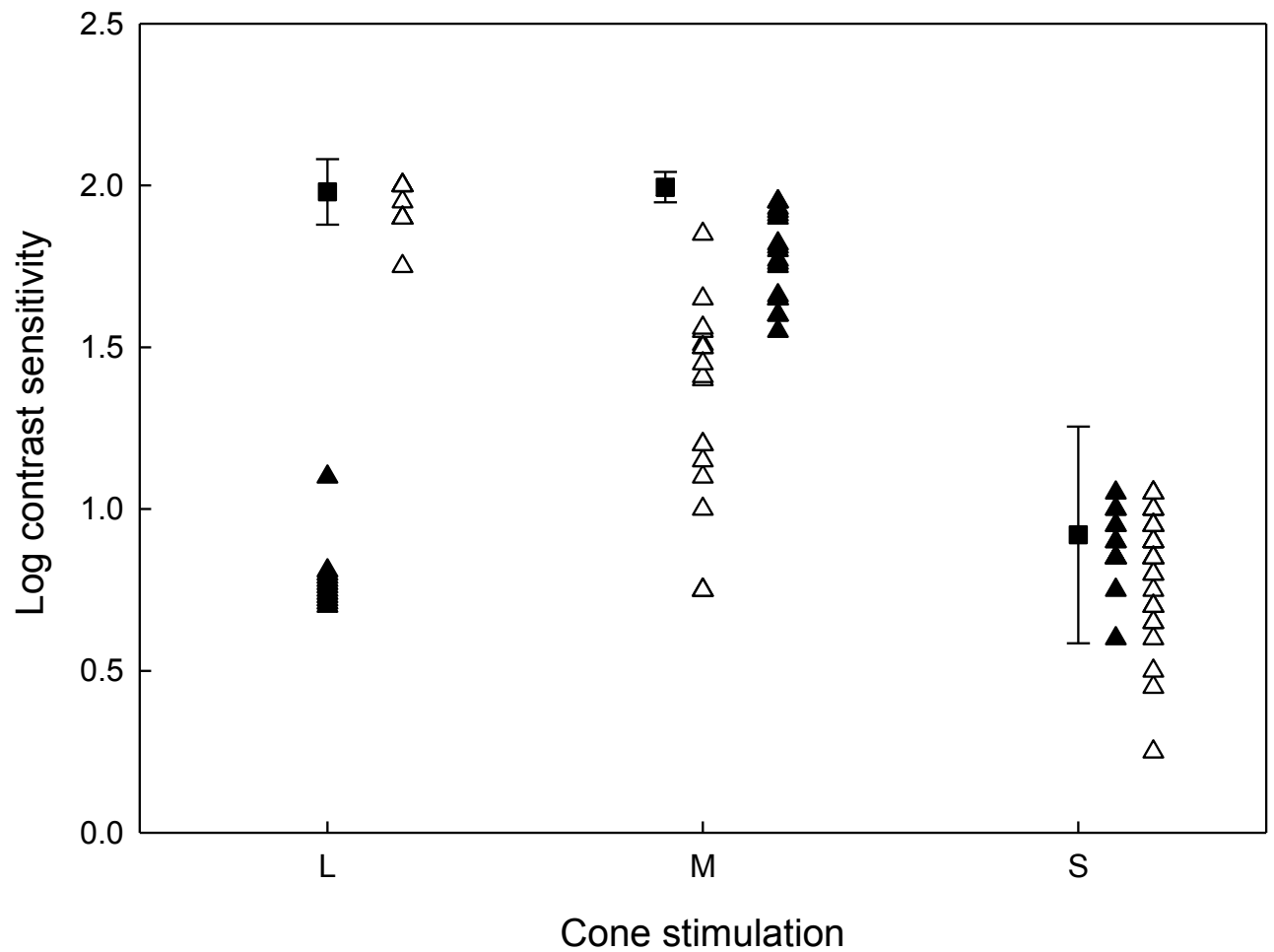
Figure 21. Log cone contrast sensitivity is plotted as a function of the L, M, and S cone stimulation. Black squares indicate the mean for normals and the error bars is ±3SD from the mean. Solid triangles are the results from individual protan subjects and the open triangles are results from individual deutan subjects.

Figures 22 and 23 show the ROC curves for protan and deutan subjects respectively. Both figures show a perfect classification to the type of defect with the maximum area under the curve for the corresponding cone contrast (i.e. L cone for protans, and M cone for

deutans). The L-cone contrast cut-off point was 1.45 and this produced sensitivity and specificity values of 100%. For the M-cone contrast, the cut-off point was higher at 1.9. This produced sensitivity for deutan defects of 97% with 100% specificity.

In terms of the area under the curves for protans, the L cone contrast has an optimum area of 1.00 whereas the M and S cone were closer to the chance line. The L and M cone contrast curves were significantly different from the chance line ($P<0.0001$, $P=0.04$ respectively), whereas the S cone contrast was not significantly different ($P=0.25$). For the deutans, the area under M cone contrast curve was equal to 1.00 and close to one for the L cone contrast. However, the area under the curve for the S cone contrast was closer to the chance line with a value of 0.67. The L, M, and S cone contrast curves were significantly different from the chance line with $P<0.001$, $P< 0.001$, and $P=0.006$ respectively. Because ROC curves show that both the L and M cone contrast sensitivities were affected in both the protans and deutans, we examined the data to ensure that the lowest abnormal cone contrast sensitivity value between the M and the L cone values classified the defect correctly. All the subjects who failed the test were classified correctly.
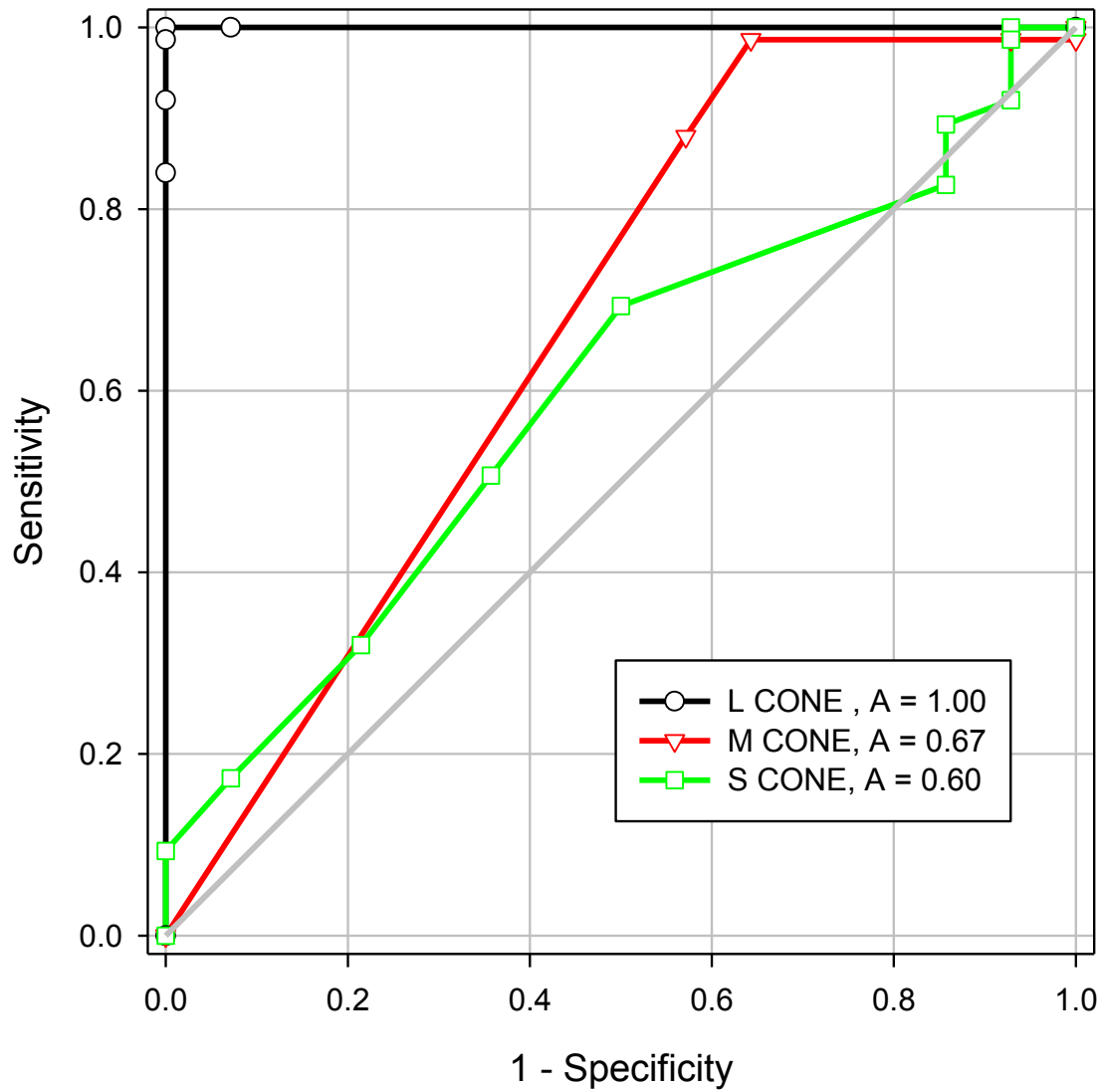
Figure 22. Receiver Operating Characteristic (ROC curve) for the protans. A indicates the area under the curve.
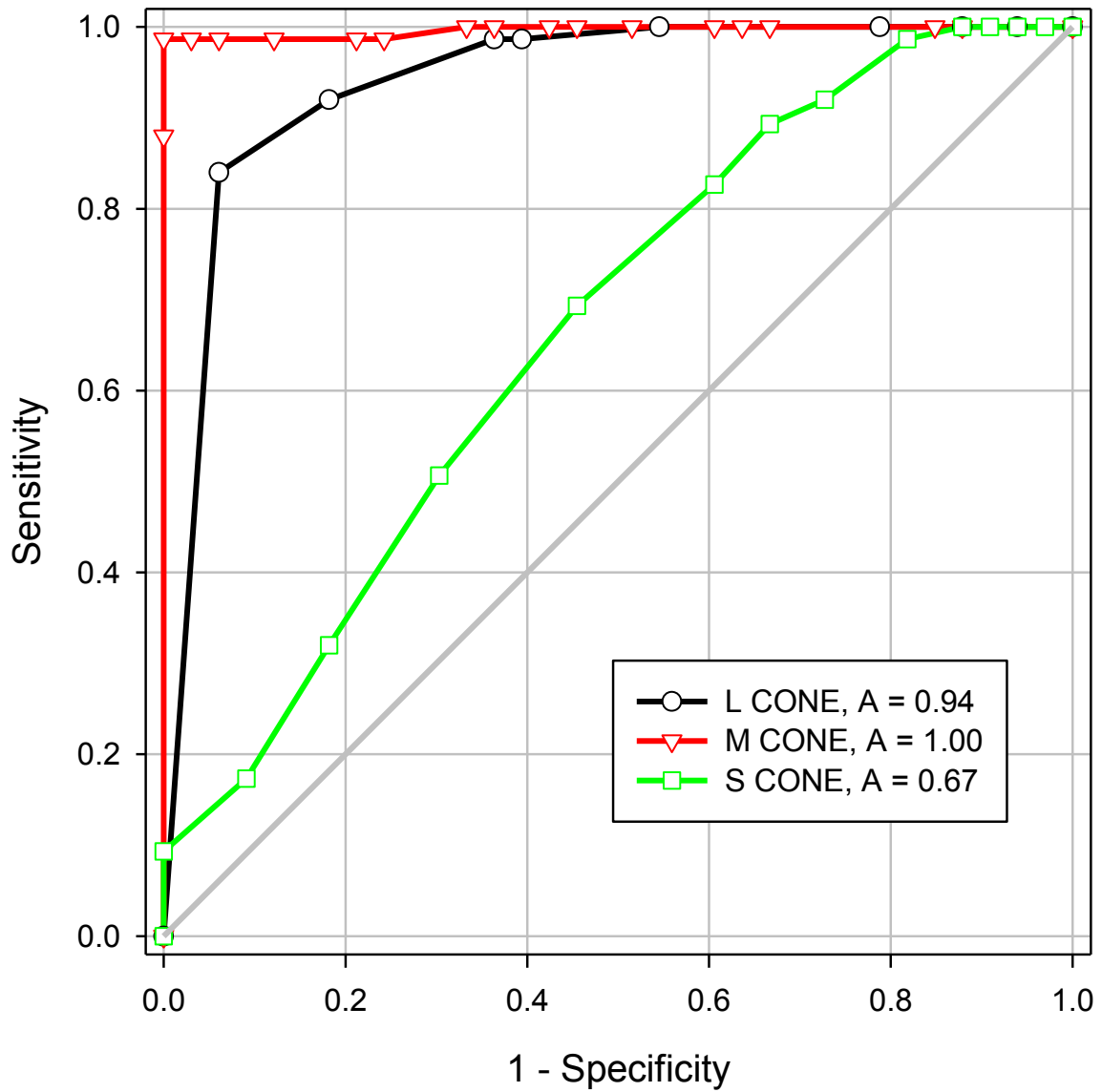
Figure 23. Receiver Operating Characteristic (ROC curve) for the deutans. A indicates the area under the curve.

Although there was near perfect agreement with the cone contrast sensitivities and the nature of the colour vision defect, having separate cut-off values for the L and M cone contrast sensitivities could be undesirable in a clinical setting since these two cone types are often assumed to be functionally similar to each other except for their spectral sensitivities. It is well established that the S cone is neurophysiologically and anatomically distinct and so the different cut-off for the S cone is not surprising (Kaiser & Boynton, 1996). We examined a number of cut-off criteria between 1.45 and 1.9 for the L and M cone contrast and the single values that was closest the separate cut-off values was 1.75 for both the L and M cones. Failure of CCST occurred when either the L or M sensitivity was less than 1.75. Table 24 shows the comparison with anomaloscope. There was perfect agreement with anomaloscope with a κ coefficient of 1.00.

Table 24. Agreement of the cone contrast test with anomaloscope for screening with 1.75 cut-off point for both the L cone and M cone contrast sensitivity

1.75cut-off		L and M cone contrast	
		ANOMALOSCOPE	
		PASS	FAIL
CONE CONTRAST SENSITIVITY	PASS	75	0
	FAIL	0	47

Table 25 shows the pass/fail repeatability for the L and M cone contrast when the criterion was set to 1.75 cone contrast. There was a good agreement with κ coefficient of 0.91 (standard error ±0.04). The lower agreement between the first and second visit was because

- One of the colour-normals who passed the cone contrast test at the first session, failed it at the second session.

- Four deutanomalous (12% of the deutan subjects) who failed the cone contrast test at the first session passed it at the second session.

Table 25. Repeatability of the L and M cone contrast sensitivity for the deutans and colour normals.

1.75 Cut-off		CONE CONTRAST 1st VISIT	
		PASS	FAIL
CONE CONTRAST 2nd VISIT	PASS	75	4
	FAIL	1	42

In classifying the defect, the lowest of the L or M cone contrast sensitivity values was used to determine whether the defect protan or deutan. This resulted in one deuteranomalous subject misclassified as protan and another deuteranomalous subject unclassified because the L and M cone values was equal. Table 26 shows the classification as protan or deutan on the cone contrast test relative to the anomaloscope. There was a good agreement between the

anomaloscope and cone contrast sensitivity test with κ coefficient of 0.90 (standard error ±0.066).

Table 26. The classification of the cone contrast tests as protan or deutan relative to the anomaloscope using the lowest contrast sensitivity value to determine the type of the defect.

		Anomaloscope	
		Protan	Deutan
Cone Contrast Test	Protan	14	1
	Deutan	0	30
	Unclassified	0	1

Table 27 shows the repeatability of the cone contrast test in terms of classification of the defect. There was a perfect agreement between the first and second session with κ coefficient of 1.00. This result may appear unusual because of the two discrepancies shown in Table 26 for the first visit. The reason for the perfect agreement between the two sessions is that the two deuteranomalous individuals passed the cone contrast test at the second visit so that Table 27 represents the between-session classification agreement for individuals who failed the test at both visits.

Table 27. Repeatability of the Cone Contrast Sensitivity test.

1.75 Cut-off		CONE CONTRAST 1st VISIT	
		Protan	Deutan
CONE CONTRAST 2nd VISIT	Protan	14	0
	Deutan	0	38

5.3.3 Prototype ColorDx Computerized Colour Vision Test

5.3.3.1 Screening plates

Using the recommended criterion of 5, or more errors, on the red-green screening series (TJ Waggoner, personal communication), most of the colour defective subjects failed the red-green screening series at the first session. There were two deuteranomalous individuals who passed the first visit. These two deuteranomalous subjects and one more deuteranomalous individual passed the test at the second session. Only one colour normal subject failed the screening test at the first session and this individual was also the only colour-normal to fail at the second session.

Table 28 shows the comparison between the pColorDx computer test and Nagel anomaloscope at the first session. There was a good agreement with κ coefficient of 0.95 (standard error ±0.029). The sensitivity and specificity were 95.7% and 98.7% respectively.

Table 28. Comparison between the pColorDx test and the anomaloscope.

		Anomaloscope	
		Pass	**Fail**
pColorDx	**Pass**	74	2
	Fail	1	45
		specificity 98.7%	Sensitivity 95.7%
Kappa coefficient		0.95 standard error ±0.029	

Table 29 shows the repeatability of the red-green screening plates of the pColorDx computer test. There was good agreement with κ coefficient of 0.98 (standard error ±0.017).

The discrepancy between the two sessions was due to the one deuteranomalous individual who failed the test at the first session, but passed it at the second session.

Table 29. Repeatability of the pColorDx test.

		pColorDx 1st	
		PASS	FAIL
pColorDx 2nd	PASS	76	1
	FAIL	0	45

5.3.3.2 Diagnostic plates

Using the recommended criteria mentioned previously in the method section, the pColorDx diagnostic plates classified correctly 100% of the red-green colour defective as either protan or deutan at both sessions. That is, the κ coefficient type of red-green defect was 1.00 and so was the repeatability of the classification.

5.3.3.3 Grading the severity

Figure 24 shows the relationship between the pColorDx classification of the severity and the matching ranges of the anomaloscope. There was an increase in pColorDx severity grade with increasing anomaloscope matching ranges. The Spearman rank correlation coefficient of 0.45 was significant (p=0.0018), but only moderate.

The reasons for the less than perfect correlation were that 10% of the dichromats (5% of the deuteranopes and 37% of the protanopes) were classified as mild or moderate on the pColorDx test. In addition, there were three deuteranomalous individuals whose range of

matches on the anomaloscope were relatively small (i.e. <20) and these individuals were classified as severe by pColorDx test.
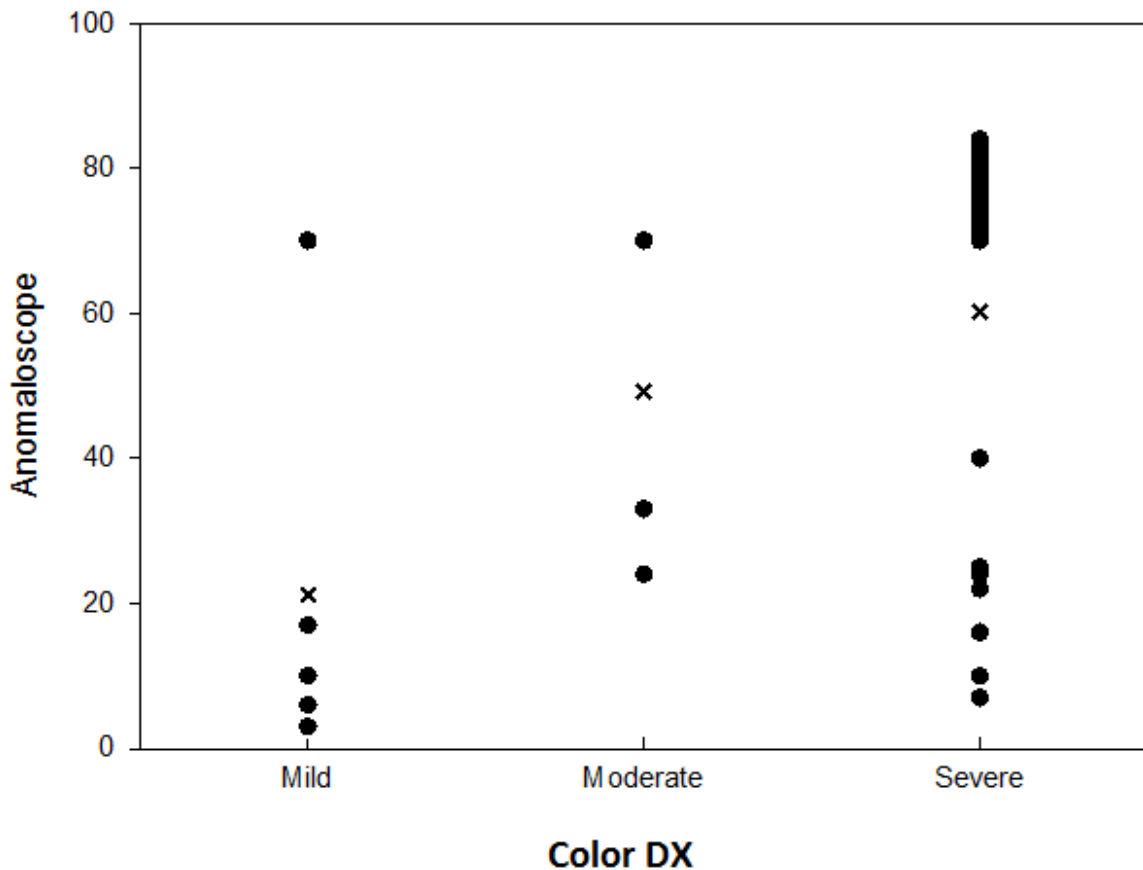
Figure 24. Relationship between the pColorDx severity classification and the anomaloscope matching ranges for colour defective subjects (dichromats individual ranges have been offset to show the number of individuals at each severity grade). X's are the average anomaloscope matching ranges for each pColorDx grade.

Table 30 shows the repeatability of the pColorDx computer test for the severity. There was a fair agreement between the first and second visit with κ coefficient close to 0.66 (standard error ±0.123). The discrepancies between the two sessions were a result of:

- 7% of the colour-defectives who were classified as severe at the first session were classified as mild or moderate at the second visit.

- 5% of the colour-defective subjects who were classified as mild at the first visit were classified as moderate at the second visit.

Table 30. Repeatability of the pColorDx computer test severity grading.

		pColorDx 1st		
		MILD	**MODERATE**	**SEVERE**
pColorDx 2nd	**MILD**	1	0	1
	MODERATE	2	4	2
	SEVERE	0	0	34

5.3.3.4 Blue Yellow Defects

In terms of the blue-yellow screening series, ten subjects (9 deutan and 1 colour normal) failed the blue-yellow screening test when the criterion was more than 2 errors on the tritan plates. Most of them passed the HRR test. Two of the individuals were the deutans who also missed some blue-yellow figures on the HRR test. Five subjects (3 deutan and the colour normal) did not show any improvement in the second session, whereas the other five showed improvement (passed the test) at the second session.

5.3.4 General discussion

All the three computerized colour vision tests are effective on screening for red-green colour vision defects. Table 31 summarizes the κ values in screening for red-green defect and the sensitivity and specificity of each test. As can be seen the cone contrast test has the highest

agreement with anomaloscope. The lower agreement between the anomaloscope and the

other two computer tests (CVRT and pColorDx) was a result of the same three

deuteranomalous subjects who passed the two computerized tests as well as one normal

subject (different subject for each test) who failed these two computerized tests.

Table 31. Sensitivity, specificity and level of agreement with anomaloscope for the three computerize tests used in this study.

	CVRT	Cone Contrast (L&M)	pColorDx
Sensitivity	92%	100%	96%
Specificity	99%	100%	99%
Kappa Coefficient	0.91	1.00	0.95
95% CI	0.84 to 0.99	1.00	0.89 to 1.00

Table 32 summarizes the κ coefficient for the repeatability for each screening test. All

the three tests were repeatable tests with perfect or close to perfect agreement between

sessions. The slightly lower agreement of the repeatability of the cone contrast sensitivity

was because

- o 1.3% of the colour normal subjects, who passed the first visit, failed the second

  visit.

- o 12% of the deutan subjects who failed the test at the first session passed it at the

  second session.

Table 32. Repeatability of the screening series for each of the computer tests.

	CVRT	Cone contrast	pColorDx
Kappa Coefficient	0.98	0.91	0.98
95% CI	0.95 to 1	0.83 to 1.00	0.95 to 1.00

Only the Cone contrast test and pColorDx computer test can classify the defect as protan or deutan. Table 33 summarizes the κ valued in terms of classifying the defect. Both tests showed that they are effective in classifying the defect as protan or deutan.

Table 33. level of agreement with anomaloscope for the cone contrast and pColorDx tests used in this study.

	Cone contrast	pColorDx
Kappa Coefficient	0.90	1.00
95% CI	0.78 to 1.00	1.00

The pColorDx test was designed to grade the severity of the defect in terms of mild, moderate to severe. The test's ability to grade the severity was moderate, but it was slightly lower than the HRR plates. The Spearman rank correlation showed that there was a statically significant but moderate correlation between the pColorDx and HRR with r=0.46 (P=0.0017). The reason for the less than ideal correlation was because

- 6% and 24% of the dichromats were classified as moderate by the pColorDx and the HRR respectively.

- One deuteranomalous with matching range <10 was classified as severe by the pColorDx but as moderate by the HRR test.

- 7 % of the colour defective subjects were classified as severe by the pColorDx but as moderate by the HRR test.

- 7% of the colour defective subjects were classified as mild by the pColorDx but as moderate by the HRR test.

All three tests can screen for blue-yellow defects. In the CVRT test, all the colour normals and colour defective subjects identified the location of the blue-green target correctly. All of their responses to the blue-green figures were less than 1.0 standard deviation from the white control value except for one deuteranomalous subject. This deuteranomalous failed the HRR and pColorDx tests at the first session. However, all blue-green z-scores were within ±1.0 of the white control at the second visit.

In the CCST, three deutans failed the S-cone contrast test with S-cone contrast values below 0.6. These three subjects passed the HRR blue-yellow screening test, but one failed the pColorDx tritan test. In terms of the repeatability, all subjects who failed the first session passed the test (had values more than 0.6) at the second session.

However, ten subjects failed the pColorDx test. Eight of them passed the HRR test without any error. At the second visit, 50% of the ten subjects (who failed at the first session) passed the second session.

One reason for missing the blue-yellow screening series may have been due to the subject expectation that they would not perceive a figure or a letter and so they did not pay attention and missed the figure. The result that most subjects passed a given tritan test at the second visit supports this hypothesis. The age of the deutan subjects might have been another reason for the blue-yellow failures. The three oldest subjects had a deutan defect and they

also failed the pColorDx tritan test at both visits. However, there were also two deutans under 25 years old who also failed the pColorDx tritan test at both visits. In addition, none of the subjects who failed the pColorDx tritan test twice failed any other tritan test twice. The higher tritan failure rate on the pColorDx test suggest that the pColorDx tritan test may be more sensitive to age related changes in colour vision more susceptible to errors of expectation, used colours that are within some deutan subjects' discrimination ellipse with gray or some combination of these factors.

5.4 Farnsworth Munsell D15 test

5.4.1 Pass/ fail

It has become common practice in Canada to administer the Farnsworth-Munsell D15 (D15) at least three times and the patient has to pass the test on 2 of the 3 trials. This practice developed because individuals taking the D15 often have to pass the test in order to qualify for an occupation and the finding that repeatability of the D15 is not perfect (Hovis et al, 2004). Although, the 2 out of 3 trials offers a method for assessing the within session variability, the repeatability of the D15 on different days using this criterion has never been assessed to our knowledge.

First, we investigated whether there is a difference in the pass/fail rates using the results from the first trial versus the 2-out-of-3 trials at the first session. The failure criterion was the Farnsworth's failure criterion of 2, or more, major crossings by visual inspection (Farnsworth, 1947).

Table 34 shows the pass/fail for just the first trial results compared with pass/fail results for 2 out of 3 trials using visual inspection. Note that only the colour defective subjects are included in the comparison because colour normals would not be taking the D15 because they would have passed the screening test. There was a good agreement between the first trial and 2 out of 3 trials with κ coefficient equal to 0.95 (standard error ±0.04). The only discrepancy was one protanope who failed the first trial, but passed the other two trials. These results indicate that there would be a learning effect for a small number of subjects when the subject performs the tests three times at one session.

Table 34. Agreement between the first trial pass/fail results and the pass/fail results for 2 out of 3 trials.

		1st Trial Only	
		PASS	**FAIL**
2 out of 3 trials	**PASS**	17	1
	FAIL	0	29

Table 35 shows the repeatability of the 2 out of 3 trials in the D 15 test. There was a good agreement with κ coefficient of 0.87 (standard error ±0.07). The discrepancies were due to 3 dichromats (2 deuteranopes and one protanope) who failed the first session, but passed at the second session.

Table 35. Repeatability of the 2 out of 3 trials in the D15 test

		1st VISIT	
		PASS	FAIL
2nd VISIT	PASS	18	3
	FAIL	0	26

5.4.2 Classification

Classification was first carried out using visual inspection for those subjects who failed on 2 out of 3 trials at the first session. Visual inspection for 2-out-of-three trials was carried out by third person who was familiar with D15, but was masked as to the anomaloscope results when carrying the inspection. There were two deuteranopes who were classified incorrectly as protan at the first visit. All others were classified correctly which resulted in a κ coefficient of 0.83 (standard error ±0.11). Table 36 shows the repeatability of the D 15 classification based on the visual inspection. There was a good agreement between the first and second visit with κ coefficient of 0.92 (standard error ±0.081). In the second visit, one previously mentioned deuteranope was now classified correctly as deutan whereas the other deuteranope was classified incorrectly as protan.

Table 36. Repeatability of the classification based on visual inspection of the D15 test.

		1st VISIT	
		PROTAN	DEUTAN
2nd VISIT	PROTAN	8	0
	DEUTAN	1	20

If Colour Difference Vector angle of ≤-3.0 degrees is deutan and >-3.0 is a protan, all the colour defective subjects were correctly classified. Moreover, the repeatability of classifying the type of the defect based on the Colour Difference Vector angle shows that there was a perfect agreement between the first and second visit.

5.4.3 Severity

The D15 only provides a classification of mild versus moderate-to-severe. Nevertheless, we were interested in the relationship between the percentage of colour defective subjects who either passed, or failed, the D15 and the severity rating on the HRR and pColorDx tests. Figure25 shows the percentage of subjects who failed the D 15 and had one of three severity grades on HRR and pColorDx test. As expected, the majority of individuals who failed the D15 were classified as having moderate to severe classification on the other two tests. However, 3.3% of the subjects who failed the D15 were classified as mild by the pColorDx.

Figure 26 shows the percentage of subjects who passed the D15 at each severity grade for the two tests. Twenty percent and 35% of the colour defective subjects who passed the D15 were classified as moderate by the HRR and pColorDx tests respectively. Moreover, 47% and 53% of the colour defective subjects who passed the D15 were classified as severe

by the HRR and pColorDx tests respectively, which was not what we expected. One possible reason for this unexpected finding was that the average colour differences between the test figure and background for the HRR and pColorDx test were smaller than in the D15. The percentage of colour defective subjects who passed the D15 and classified as moderate in this study was close to what Cole et al (2006) found. However, the percentage of the colour defective subjects who passed the D15 and were classified as severe by the HRR in this study was higher than reported by Cole, et al (2006). The possible reason for the different finding was that they had more subjects with a mild defect. Note that all the colour defective subjects who passed the D15 were anomalous trichromats except one protanope who was classified as severe in both HRR and pColorDx tests.
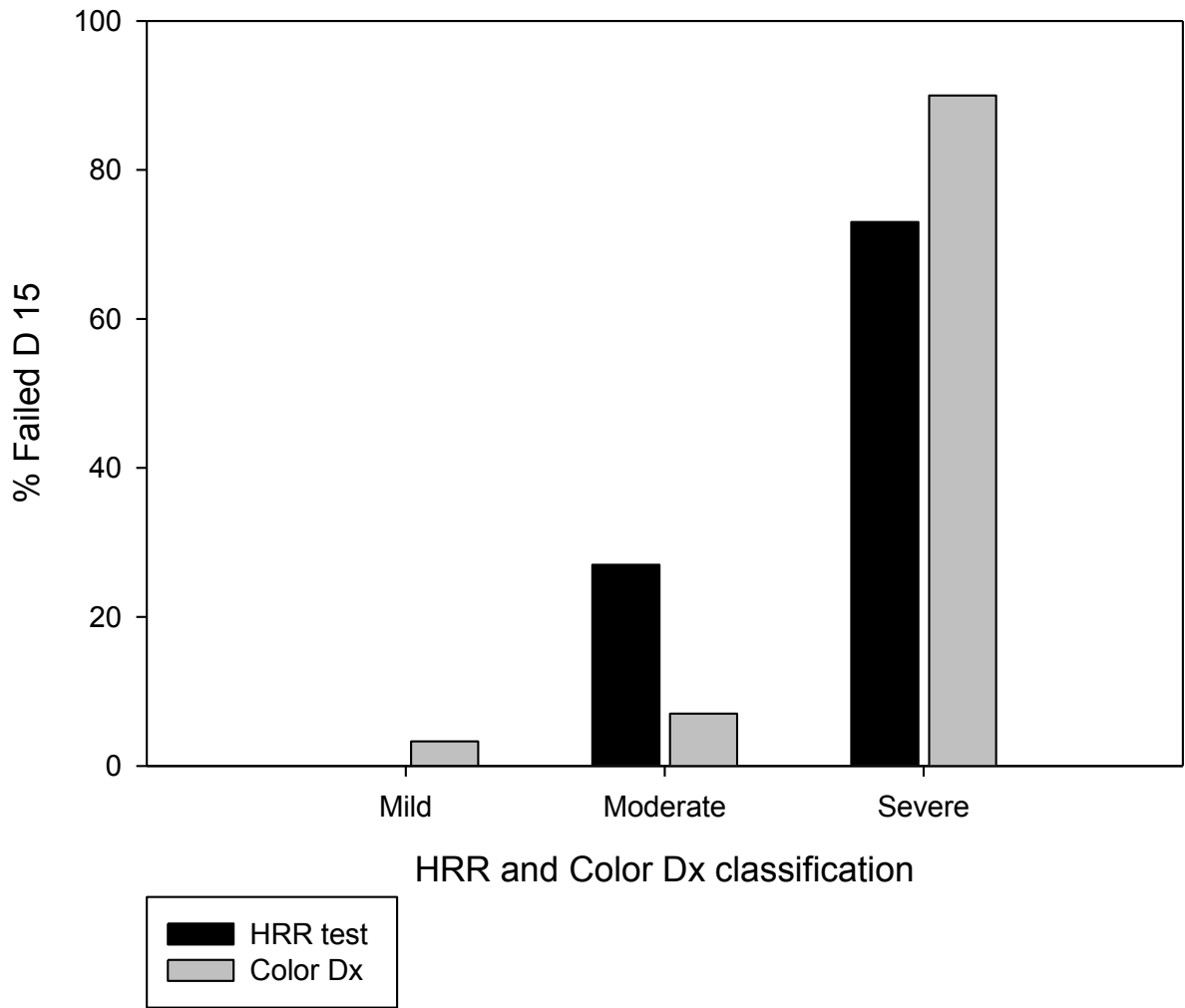
Figure 25. Percent of subject who failed the D15 and were classified as mild, moderate, or severe in the HRR and pColorDx computer tests.
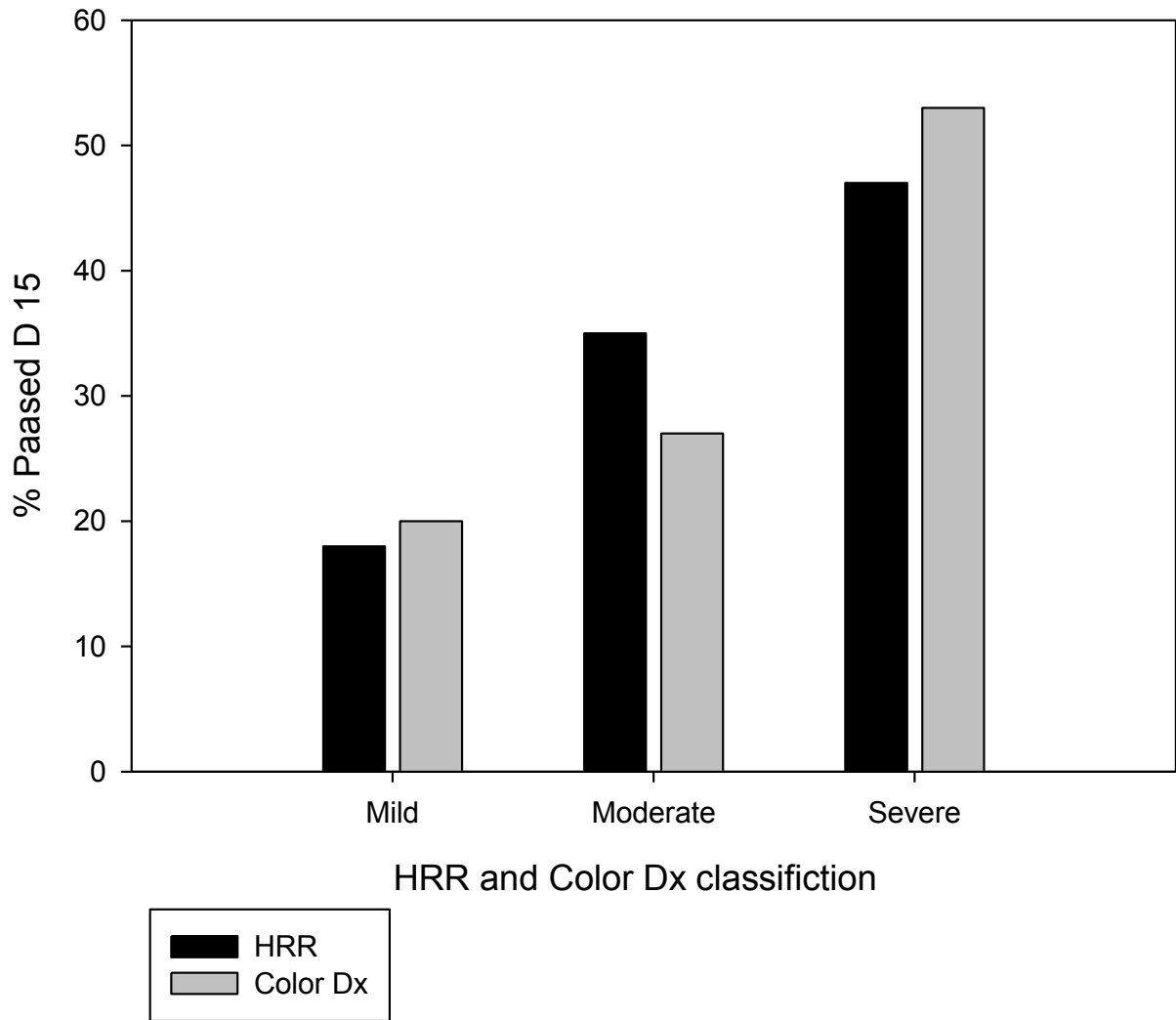
Figure 26. Percent of subject who passed the D15 and were classified as mild, moderate, or severe in the HRR and pColorDx computer tests.

5.4.4 General discussion

All the dichromats and one deuteranomalous failed the D15 test. This deuteranomalous had a severe defect in the anomaloscope (> 40 matching range). Our study confirmed that there is a small learning effect on the D 15 when the test is performed multiple times within a session. One subject who failed at the first trial showed improvement in the second and third trial. This would support the previous finding in the literature that allowing the subject to review the first arrangement could increase the pass rate (Hovis et al 2004; Birch, 2008).

Sixty four percent of subjects with red-green defect failed the test using the fail criterion of two, or more, major crossings ($\geq$1.7 C-index) on the 2-out-of-3 trials. The repeatability of the D15 (2 out of 3 trials at first vs second visit) was good with κ coefficient of 0.87 (standard error $\pm$0.07). The repeatability of the D15 in this study was slightly more than the value calculated by Hovis et al (κ =0.84) who had the subjects perform only one trial at each session, but less than the value calculated from Farnsworth data (κ =0.96) who included a large number of colour normals in his study which improved the overall repeatability. Although there was only one subject who improved to a pass using the 2-out-of-3 rule, this could have a major impact on this person's career if he had to pass the D15 in order to qualify for a job. Interestingly this person also failed the first trial at the second session, but passed using the 2 out of 3 rule. Note that none of the subjects in this study passed on the first trial and then failed based on the 2 out of 3 rule.

The differences in the average Colour Difference Vector parameters (Angle, S-index, and C-index) were determined for the first and second visit for 2 out of 3 trials and were calculated to determine whether they fell within the range defined by the coefficient of

repeatability published previously (Hovis et al., 2004). Table 37 shows the percent of

subjects who had values within those ranges reported by Hovis et al. In general, the results

were most similar for the C-index and the S-index with the angle having most subjects

outside of the coefficient of repeatability. This last result is expected because the perfect

arrangement results in an angular value of 45 degrees and if there was a perfect arrangement

on one trial, but only a single crossing or transposition in the second trial, then the difference

in the angles could be close to 45 degrees. If there was one or two crossings at one session

and just one or two transposition at another session, then the difference in angles could

approach 90 degrees.

Table 37. Percent of subjects who had values (different in parameters between the first and second visit) within the coefficient of repeatability values calculated by Hovis et al (2004)

	Angle	C-index	S-index	Crossing
Deuteranomalous	100%	92%	92%	92%
Deuteranope	81%	90.5%	90.5%	100%
Protanomalous	80%	100%	100%	100%
Protanope	87.5%	100%	100%	100%

The repeatability of the classification based on visual inspection agreed on about 97%

of the subjects. This value was better than the 80% repeatability of classification in the

Hovis, et al study (2004). This was likely a result that the current study had a higher

percentage of dichromats relative to the earlier study. The dichromats were more likely to

multiple crossings and judging the orientation of the several crossings may be easier than judging the orientation of just two or three crossings.

The type of the defect identified by the D15 Colour Difference Vector angle has perfect agreement with both HRR and pColorDx computer test classification. However, the agreement based on the visual inspection of the D15 results with both HRR and pColorDx computer were not as good with κ coefficient of 0.84 (standard error ±0.10). Table 38 summarizes the agreement of the classification between the visual inspection of the D15 arrangements with both HRR and pColorDx. The lower agreement between the D15 with both HRR and pColorDx tests was due to two deuteranopes who were classified as protan with the D15 results.

Table 38. Classification agreement between the D15, HRR and the pColorDx computer.

		HRR		pColorDx	
		Protan	Deutan	Protan	Deutan
D 15	Protan	8	0	8	0
	Deutan	2	20	2	20

Chapter 6
Summary and Conclusion

6.1 Pseudoisochromatic plates

The three pseudoisochromatic plate tests (HRR, Ishihara, and PIPC) can be used confidently to detect red-green colour vision defects. Using the anomaloscope results as the validation criterion, all the three tests show a very good to perfect agreement with anomaloscope in screening for the red-green colour vision defects. The agreement with anomaloscope for screening purposes was slightly lower for Ishihara and PIPC tests than the HRR test. The Ishihara and PIPC tended to classify mild deuteranomalous individuals as colour-normal. However, the differences in the agreement with anomaloscope were not statically significant. In terms of the repeatability of the red-green screening plates, the study confirms that the three tests are highly repeatable tests.

Both HRR and PIPC tests can screen for blue-yellow defect. There were 2 deuteranomalous at the first visit and a different deuteranomalous at the second visit who made single error in the HRR test. In the PIPC test, only one deuteranomalous subject failed the blue-yellow screening plates at the first visit with two errors. This deuteranomalous was one of the individuals who made errors in the HRR test at the first visit. At the second visit, all the subjects (normals and colour defectives) passed the blue-yellow plates on the PIPC test.

The study showed that of the three pseudoisochromatic tests, the HRR is the most effective in classifying the defect as either protan or deutan. It had perfect agreement with anomaloscope in terms of the type of the defect if the subject failed the screening test. The

92

difference between the HRR and the other two tests was statically significant. One possible reason for the better performance of the HRR was that it presented more diagnostic plates compared with either the Ishihara or PIPC tests. Another reason for lower agreement with Ishihara was that our procedure only recorded missed figures. We did not ask the subjects which figure was more visible if they reported with both figures. It is possible that the agreement with anomaloscope would increase if this question was asked.

In terms of the repeatability of the classification plates as either protan or deutan, the study confirms that the HRR test is highly repeatable test followed by the Ishihara test (good repeatability) whereas the repeatability of the classification plates was reasonable on the PIPC test. Again, this was probably due to the large number of diagnostic plates presented by the HRR test.

The HRR test is the only one of the three tests that can determine the severity of the defect as mild, moderate, and severe. The HRR was able to separate the mild from the moderate and severe defect, but it was not as effective in separating moderate from severe defects. This result suggests that the HRR should not be used alone in determining the severity of the defect. Other tests, such as the Farnsworth D15 test or the anomaloscope, should be administered if determining the severity of the defect is important for occupational reasons.

6.2 Computerized tests

The study has shown that the new test (CVRT), Cone Contrast Sensitivity test, and pColorDx test were effective in terms of screening for red-green defects and can be as effective as the printed tests evaluated in this study. The Cone Contrast test has the highest agreement with

the anomaloscope, but it is not significantly different from the other two tests. All the three tests were understood and accepted by the adult subjects. All the three tests were highly repeatable tests.

All the three tests were capable of screening for blue-yellow defects. In the CVRT test, all but one deuteranomalous subject identified the location of the blue-green target correctly within 1.0 standard deviation from the white control. This deuteranomalous failed the blue-yellow plates on the HRR and pColorDx tests at the first visit. In the pColorDx test, ten subjects (nine deutan and one colour normal) failed blue-yellow series. Eight of them passed the blue-yellow screening plates on the HRR test without any error. However, three subjects failed the cone contrast test with cone contrast below 0.6. All the three passed the tritan plates on HRR test. At the second visit, all the subjects (normals and colour defectives) passed the cone contrast tritan test.

The Cone contrast and pColorDx test were able to classify the defect as protan or deutan. Both tests showed very good to perfect agreement with anomaloscope in terms of classifying the defect as protan or deutan.

The ability of the pColorDx test to grade the severity of the defect was moderate, and slightly lower than the HRR plates. The correlation between the severity results of the pColorDx and HRR was moderate. Although comparable to the HRR, the pColorDx test took three times longer than the HRR test to administer.

6.3 Farnsworth D15

The study showed that there was small learning effect on the D15 when the test was performed multiple times in one session. The test has reasonable pass/fail repeatability when

we used the 2 out of 3 rule. This was a marginally better repeatability than using the results from only one trial on separate days. The study confirmed that the Colour Difference Vector parameters were slightly more effective criteria in classifying the type of the defect. In terms of the repeatability of classification, the study showed that there was a good agreement between sessions based on the visual inspection and perfect agreement between sessions based on Colour Difference Vector parameters.

Bibliography

Bailey, James E., Neitz, Maureen., Tait, Diane M., Neitz, Jay. "Evaluation of an updated HRR color vision test." *Visual Neuroscience* 21.03 (2004): 431-436.

Birch, J. (1993). *Diagnosis of defective colour vision* Oxford University Press Oxford, United Kingdom.

Birch, J. (1997). Efficiency of the Ishihara test for identifying red-green colour deficiency. *Ophthalmic and Physiological Optics, 17*(5), 403-408.

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

Cole, B. L., Lian, K., & Lakkis, C. (2006). The new Richmond HRR pseudoisochromatic test for colour vision is better than the Ishihara test. *Clinical and Experimental Optometry, 89*(2), 73-80.

Dain, S. J. (2004). Clinical colour vision tests. *Clinical and Experimental Optometry, 87*(4-5), 276-293.

Deeb, S. S. (2004). Molecular genetics of colour vision deficiencies. *Clinical and Experimental Optometry, 87*(4-5), 224-229.

Farnsworth, D. (1947). Farnsworth dichotomous test for color blindness. *San Antonio, TX, US: Psychological Corporation.*

Hovis, J.K, Ramaswamy, S., & Anderson, M. (2004). Repeatability indices for the Farnsworth D-15 test. *Visual Neuroscience, 21*(03), 449-453.

Kaiser, P. K., & Boynton, R. M. (1996). Human color vision. Optical Society of America Washington, DC.

Lakowski, R. (1966). A critical evaluation of colour vision tests. *The British Journal of Physiological Optics, 23*(3), 186-209.

Lakowski, R. (1969a). Theory and practice of colour vision testing: A review. 1. *British Journal of Industrial Medicine, 26*(3), 173-189.

Lakowski, R. (1969b). Theory and practice of colour vision testing: A review: 2. *British Journal of Industrial Medicine, 26*(4), 265-288.

Ling, B. Y., & Dain, S. J. (2008). Color vision in children and the Lanthony new color test. *Visual Neuroscience, 25*(03), 441-444.

Pokorny, J., Smith, V., Verriest, V.,& Pinckers, A. (1979). *Congenital and acquired color vision defects* Grune & Stratton New York.

Procedures for testing color vision: Report of working group 41(1981). .Washington, D.C. : National Academy Press.

Rabin, J. (1996). Cone-specific measures of human color vision. *Investigative Ophthalmology & Visual Science, 37*(13), 2771-2774.

Rabin, J. (2004). Quantification of color vision with cone contrast sensitivity. *Visual Neuroscience, 21*(03), 483-485.

Schneck, M. E., Haegerstrom-Portnoy, G., Lott, L. A., & Brabyn, J. A. (2014). Comparison of panel D-15 tests in a large older population. *Optometry and Vision Science, 91*(3), 284-290.

Shin, Y. J., Park, K. H., Hwang, J., Wee, W. R., & Lee, J. H. (2007). A new color vision test to differentiate congenital and acquired color vision defects. *Ophthalmology, 114*(7), 1341-1347.

Steward, Judy, M, and Barry, L. Cole. "What do color vision defectives say about everyday tasks?." *Optometry & Vision Science* 66.5 (1989): 288-295.

Tiffin, J., & Kuhn, H. S. (1942). Color discrimination in industry. *Archives of Ophthalmology, 28*(5), 851.

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.

Waggoner, T. (2003). Ishihara compatible pseudoisochromatic plate (PIPIC) color vision test. *T.L Waggoner, Inc. Gulf Breeze, FL.*

Appendix A
Questionnaire

1) **Other than wearing spectacles or contact lenses and/ or having colour vision problem, are you aware of any vision problems or problems with the eyes, or are you being treated for any eye diseases?**

    Yes….   No….

 Eligible conditions are (circle):

History of red or pink eye, dry eye, post cataract surgery, strabismus, amblyopia, bump on

lids, past contact lens complications.

All other diseases or problems are disqualifying.

2) **Do you have a colour vision defect?**

    Yes…...   No……   Unsure………

    If yes at age did you first time become aware of the problem with colour

    judgments.

3) **Did your colour vision problem affect your career choice?**

    Yes….   No….

    If yes, were you made aware by someone that it could affect your career?

    Yes….   No….

    If yes, who informed you?