AOSIS

The Cambridge Colour Test: Reliability of discrimination trivectors in colour space

CrossMark
click for updates

Authors:
Nabeela Hasrod[1]
Alan Rubin[1]

Affiliations:
[1]Department of Optometry, University of Johannesburg, Doornfontein, South Africa

Corresponding author:
Nabeela Hasrod, nabzicakes@gmail.com

Background: The reliable assessment of colour discrimination has become increasingly important as some ophthalmic diseases and systemic conditions manifest themselves via acquired impairments of colour vision. The Cambridge Colour Test (CCT) is a computerised procedure, developed using the basic principles of traditional pseudoisochromatic plates to evaluate colour function or discrimination.

Aim: The study investigated reliability of the CCT, with emphasis on the Trivector subtest, which measures thresholds along the three protan, deutan and tritan confusion lines to probe the sensitivity of the long, medium and short wavelength cones.

Setting: The study used a prospective observational and quantitative design and took place in a research and clinical environment within the Department of Optometry at the University of Johannesburg, South Africa. Participants were 20 young adults of both genders aged 19–24 years with normal colour vision.

Methods: Two measurements (test and retest) of the Trivector test were measured monocularly (right eye) for all participants to assess the reliability of repeated measurements. Reliability was assessed using coefficients of repeatability, coefficients of variation and Bland–Altman plots with limits of agreement (LoA).

Results: Outcomes for the Bland–Altman LoA showed good intra-individual agreement and coefficients of repeatability and reliability revealed generally high test–retest repeatability. Multivariate stereo-pair scatter plot comparisons of means and differences of test and retest measurements show that points cluster tightly within the same region in the three-dimensional colour space, and the centroids of the ellipsoids representing the mean differences are close to the origin at zero.

Conclusion: Although departure from normality and outlying values can have important influences, essentially the CCT provides very similar test and retest results for the Trivector test.

Introduction

The assessment of colour vision of an individual is an essential element of routine eye and vision examinations. Colour vision deficiencies are relatively common and defects of colour vision can be detected through a variety of tests. Colour vision and its deficiencies can be assessed qualitatively and quantitatively using tests that are grouped as arrangement tests, matching tests, vocational tests or pseudoisochromatic plates.[1,2,3,4,5,6,7,8]
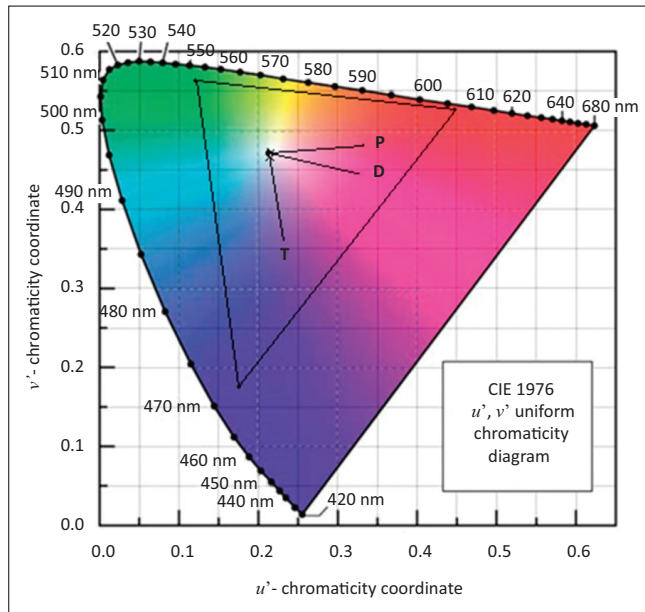
Specifically, a more sophisticated computerised procedure, the Cambridge Colour Test (CCT) (Cambridge Research Systems Ltd. [CRS], Rochester, United Kingdom),[9] a diagnostic tool broadly used in both research and clinical settings, was developed using the basic principles of traditional pseudoisochromatic plates and modern technology of randomised luminance to evaluate chromatic sensitivity or discrimination based only on chromatic cues for detecting a target or pattern (a Landolt C in the case of the CCT). Unlike previous computerised colour vision tests, an advantage of this (CCT) computer-controlled test is the ability to adjust test difficulty depending on the patient's performance, as well as random presentation of plates to reduce memorisation effects.[9,10]

The CCT uses a computer version of pseudoisochromatic plates with two testing procedures, the Trivector subtest and the Ellipses subtest.[10,11,12] The Ellipses test measures colour discrimination or sensitivity along 8, 12, 16 or 20 vectors or directions within a colour plane in the colour space (the *Commission Internationale de l'Eclairage* [the International Commission on Illumination] – CIE (1976) *u'v'* space) to produce three discrimination or MacAdam ellipses that lie along the same

Source: Adapted from Shubert EF. Colorimetry. In: Light-emitting diodes, 2nd ed. Cambridge: Cambridge University Press; 2006.

CCT, Cambridge Colour Test; CIE, International Commission on Illumination (1976) $u'v'$ space.

FIGURE 1: Three confusion lines (protan [P]; deutan [D]; tritan [T]) are indicated in the triangular 1976 $u'v'$ colour plane (the colour space itself is three-dimensional). In the CCT Trivector test, sensitivity is measured along three confusion lines and the less sensitive an individual is to a specific wavelength, the further away the resultant measurement (indicated by a cross) will be from the co-punctal point (and hence the larger the value). In this figure, the three crosses overlap and are very close to the co-punctal point.

tritan confusion line.[10,12] The Trivector test, the focus of this article, is a short (3 min – 4 min) test that measures thresholds along the three confusion lines (see Figure 1)[13] to probe the sensitivity of the long, medium and short wavelength cones.[9] Each target or stimulus to the patient or participant differs from the background along one of the three lines in the colour space, that is, of either the deutan, protan or tritan confusion lines. The three confusion lines converge to a point called the co-punctal point. Mollon (2000)[9] explains that dichromacy of colour discrimination is the basic principle underlying the CCT. A pair of physical stimuli can be chosen that yield the same chromaticity for two of the three classes of cones but differ with regard to the remaining class. In alternating such stimuli, one can establish the integrity of the one isolated class of cones. For instance, consider a point that lies in the plane of the chromaticity of the medium wavelength and long wavelength cones. If a line were to pass orthogonally through that point, lights along that line would vary only in the line of the chromaticity of the short wavelength cones. This is called the 'tritanope confusion line', as someone who lacks short wavelength cones will confuse chromaticities in that line or direction. A tritanope, deuteranope or protanope is a person who confuses chromaticities along the short, medium or long wavelength confusion lines, respectively.[9] Such confusion lines are shown in Figure 1.[13] Sensitivity is measured along these confusion lines (or vectors) and the less sensitive an individual is to a specific wavelength, the further away the resultant measurement will be from the co-punctal point (and hence the larger the value). Expected values for the CCT Trivector test are 100×10^{-4} $u'v'$ units for

both the protan and deutan lines and 150×10^{-4} $u'v'$ units for the tritan confusion line.[9]

The CCT has been used in a variety of clinical studies[14,15,16,17,18,19] and examples include the determination of colour discrimination losses in patients with ageing[14,15] or possibly chronic diseases such as multiple sclerosis[17] and for differential diagnostics of relative damage to chromatic pathways in conditions such as Parkinson's disease.[18] The CCT has also been adapted for use with animals such as squirrel monkeys, and the adapted version is said to be more reliable than other more conventional methods, because the stimulus parameters could be adjusted so that the animals were not able to use luminance differences to make correct discriminations.[19]

This article concentrates on young adult observers with colour-normal vision to assess short-term test and retest repeatability and reliability of the CCT Trivector subtest. (Refer to the work by Hasrod[20] for reliability and repeatability of the CCT Ellipses subtest.)

Research objectives

The primary aim of this study was to investigate in (colour-normal) human eyes the short-term test and retest reliability and reproducibility of results produced by the Trivector subtest of the CCT, using both univariate and multivariate statistical analyses. As such it may contribute or encourage greater use of the CCT in the future as a clinical or research tool to assess colour discrimination. Research on reliability in this study using multivariate statistics and methods was applied[20] to CCT data, and some of these methods may possibly provide an easier way of analysing the trivariate (protan, deutan and tritan) vectors as unitary vectors (or points) in colour 3-space. This article demonstrates some of these methods.

Methodology and design
Study design and setting

The study used a prospective observational and quantitative design and took place in a research and clinical environment within the Department of Optometry at the University of Johannesburg, South Africa.

Study population

The study conformed to the tenets of the 2013 Declaration of Helsinki[21] and was approved by the Higher Degrees and Ethics Committees of the university. The adult participants were predominantly of Caucasian (white) descent (60%) and 65% (or 13 of 20) were female. Mean age and standard deviation (s.d.) were 20.55 ± 1.43 years, the age range was from 19 to 24 years and the median age and quartile deviation (QD) were 20.00 ± 0.5 years. The mean auto-refraction of the 20 right eyes was $-0.54–0.20 \times 11$. (Matrix methods were used to determine this mean.) Preliminary colour vision tests (the Farnsworth D15 test and the

computerised Farnsworth–Munsell 100-hue test) were used to select participants with apparently normal colour vision. In addition, each participant's right eye, to be tested using the CCT, underwent screening (biomicroscopy, ophthalmoscopy and refractive status; distant and proximal visual acuity) to estimate the appropriate habitual or spectacle compensation. Inclusion criteria were normal colour vision and a best compensated distance visual acuity of 6/9 or better. Exclusion criteria were factors that might influence colour vision adversely: history of congenital or acquired colour vision deficiencies, ocular disease, neurological or systemic diseases or use of medication known to affect colour vision.

Materials and data collection

Cambridge Research Systems Ltd. provided the necessary hardware and software for calibration and implementation procedures (ColorCAL colorimeter; calibration software and CRS/VSG 2/5 graphics card). Stimuli were displayed on a high resolution 22″ monitor (HP CRT p1230) that was gamma-corrected. The CCT stimuli were computer-generated pseudoisochromatic images with each containing a Landolt C, as described, against a background comprised of circles of varying size, colour and luminance. A field or background of a single chromaticity or colour such as grey is used, but it is made up of smaller and larger discs that vary in terms of their luminances, which are randomised. The CCT test stimuli backgrounds are designed to eliminate luminance or contour cues; the figures (the test stimuli of Landolt letters) must be identified solely by their hues,[15] and correct identification of the gaps or openings in the stimuli are used to confirm observer awareness of the stimuli themselves. Participants use a four-option response box to indicate openings.

A pilot study was conducted[20] with two young adult participants to ascertain the duration of a session while employing both the Trivector and the Ellipses subtests. An experimental session for both tests took approximately 90 min and two sessions, for test and retest, were run either on the same day, separated by a short resting period, or on two consecutive days. However, most of the time was for the Ellipses test. Because of the time constraints and the actual testing time per test being between 20 min and 60 min, it was decided on the basis of the pilot study to measure[20] only the 8-vector and 20-vector ellipses twice each per participant in

conjunction with the shorter Trivector test (which took approximately 3 min – 4 min per participant), which also would be done twice per participant. Here only data for the Trivector test will be included. Two measurements (test and retest) of the Trivector test were measured for all participants ($n = 20$) to assess the reliability of test and retest measurements.

Data analyses

Repeatability and reliability were assessed using hypothesis tests, intraclass correlation coefficients, coefficients of repeatability (CR), coefficients of variation (CV) and Bland–Altman plots with limits of agreement (LoA). The Bland–Altman plot may be used to assess the repeatability of a method by comparing repeated measurements using one single method on a series of subjects. The graph can also be used to check whether the variability or precision of a method is related to the size of the characteristic being measured.[22] Furthermore, multivariate analysis by means of stereo-pair scatter plots was used to graphically demonstrate the short-term test and retest variation of results produced by the Trivector subtest.

Ethical considerations

The study conformed to the tenets of the 2013 Declaration of Helsinki and was approved (HDC01-24-2014 and AEC01-24-2014) by the Higher Degrees and Ethics Committees of the Faculty of Health Sciences, University of Johannesburg, South Africa.

Results

Hypothesis tests for normality in Table 1 formally test whether the population that the sample represents is likely to be normally distributed. The null hypothesis (H_0) states that the sample (and population) is normally distributed, against the alternative hypothesis (H_a) that it is not normally distributed. Shapiro–Wilk tests were utilised in conjunction with skewness and kurtosis coefficients in order to establish whether the Trivector test and retest measurements were well-modelled by normal distributions. The Shapiro–Wilk values (W) in Table 1 are between 0 and 1, with smaller values of W suggesting a rejection of normality and a value of 1 indicating normality of data. If the p-value is less than the chosen alpha level, namely 0.05, then the null hypothesis is rejected.[23]

Thus, the protan vector's test and retest samples and the deutan vector's retest sample were drawn from normally distributed populations. However, the other three samples

TABLE 1: Descriptive statistics.

Variables	Means	Medians	IQR	s.d.	Skewness (γ)	Kurtosis (κ)	Shapiro–Wilk	
							W-value	p
Protan (test)	55.40	52.00	17.50	14.70	0.31	−0.49	0.97	0.71
Protan (retest)	62.30	62.50	27.50	18.89	0.07	−0.47	0.98	0.90
Deutan (test)	54.45	51.50	20.00	14.44	0.83	−0.38	0.89	0.03
Deutan (retest)	61.35	63.50	24.00	17.95	0.07	−0.32	0.99	0.98
Tritan (test)	88.75	77.50	50.00	29.86	0.74	−0.81	0.89	0.02
Tritan (retest)	88.45	78.50	32.50	46.36	3.36	13.17	0.61	0.00

Note: All values in the first four columns (means, medians, IQR and s.d.) must be multiplied by 10^{-4} $u'v'$ units including the interquartile range and standard deviation. Skewness and kurtosis coefficients and Shapiro–Wilk results for the protan, deutan and tritan test and retest vectors for the Trivector test are also included. The sample consisted of 20 right eyes of young adults aged 19–24 years.
IQR, interquartile range; s.d., standard deviation.

were not normally distributed. In conclusion, three samples (in Table 1) were normally distributed and three were not normally distributed. Therefore, non-parametric statistics were mainly applied here and box-and-whisker plots (Figure 2) were used to assess sample medians and the spread of data via interquartile ranges.

The skewness coefficients (ɣ) in Table 1 for the protan and deutan vectors were all close to zero and less than the standard errors of skewness (s.e.) value of ± 1.095 (that is, $2\sqrt{\frac{6}{N}}$ or $2\sqrt{\frac{6}{20}}$), suggesting that with one exception (3.36), the data is not markedly skewed.[24,25] The tritan vector's retest skewness coefficient was 3.36 and, as with the normality plot, indicated that the data was positively right-skewed. The Pearson kurtosis index (κ) values (see Table 1) were slightly negative, suggesting minimal platykurtosis[26] of data except for the tritan vector's retest κ value of 13.17, indicating that this sample was markedly leptokurtic but an outlier had an effect on the normality of this sample.

The test and retest medians (Table 1 and Figure 2) for the vectors (P, D or T) were similar in value, suggesting little variation in test and retest medians. Looking at the spread of the data by means of their interquartile ranges (Table 1 and Figure 2), the values again were also mostly within a similar range, suggesting that relatively good consistency occurs between the test and retest responses. The very slightly larger IQR and medians for tritan vectors may suggest that results were less consistent and one or more outlying values may be relevant. However, hypothesis tests, which will be discussed shortly (in Table 5), indicated that test and retest measurements produced by vectors (p, d and t) for the Trivector test were repeatable and reliable, as their p-values were greater than the significance level of 0.05.

A multivariate analysis of the Trivector data was applied to further explore the samples. Multivariate data are used to better understand the associations or relations between multiple variables. Stereo-pair comets or trivariate discrimination vectors are used to provide a holistic and relatively simple graphical representation that joins test and retest vectors for the 20 right eyes concerned. The three axes of a 3-D (three-dimensional) graph or stereo-pair represents the protan (p), deutan (d) and tritan (t) vectors, respectively, for the eyes in the sample and each point in the stereo-pair scatter plot represents a unique combination of the protan, deutan and tritan components by the vector c_i, where

$$c_i = (p_i \, d_i \, t_i)' \text{ where } i = 1, 2, \ldots \infty. \qquad \text{[Eqn 1]}$$

The origin generally represents zero or (0   0   0)'. By plotting c on a set of three mutually orthogonal axes p, d and t, a three-dimensional scatter plot of a colour space is produced and the two halves of the scatter plot must be fused by allowing the eyes to diverge into an exo-position (fixation to an imaginary point behind the paper) to create the 3-D percept concerned.[27]

In Figure 3 a stereo-pair comet of test and retest results for each study participant can be illustrated graphically by means of joining a dot and a line segment with dots representing test vectors (e.g. $c_{test} = (0.2 \quad 0.1 \quad 0.1)'$) of the colour response and the ends of lines representing corresponding retest vectors (e.g. $c_{retest} = (0.3 \quad 0.2 \quad 0.2)'$).[20,27] Test values for the protan, deutan and tritan trivectors of a person can thus be represented via a single point and plotted within the 3-space. The same applies to the retest values for the same individual; then the comets (short line segments) can be used to link corresponding

Note: Small squares represent the medians of the respective vectors, the boxes are the interquartile ranges and the ends of the whiskers are the minimum and maximum values. Each box-and-whisker plot is based on results for 20 right eyes of adults aged 19–24 years. Units for the y-axis are CIE 1976 $u'v'$ units.
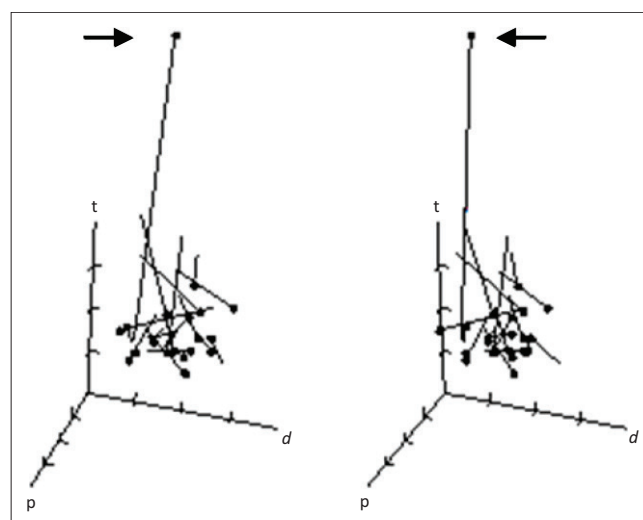CIE, International Commission on Illumination.

FIGURE 2: Box-and-whisker plots for the test (t) and retest (rt) Trivector measurements for the protan (P), deutan (D) and tritan (T) colour vectors.
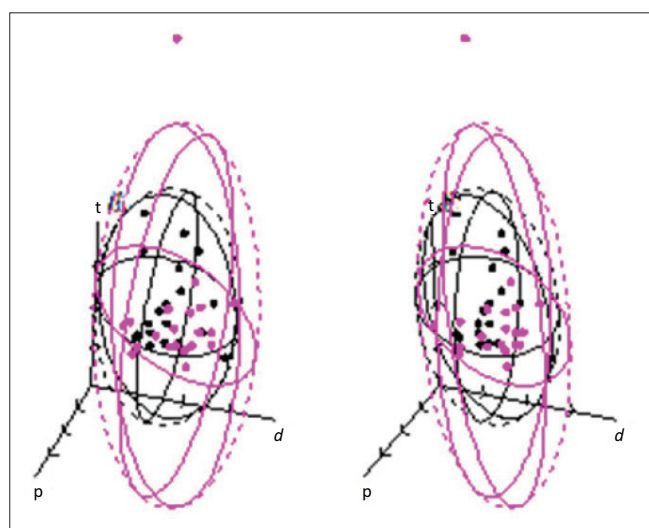

Note: Each comet joins a test measurement (represented by a dot) to the end of the line (comet), being the retest measurement for a specific eye. Vectorial combinations of the protan (p), deutan (d) and tritan (t) vectors of the CCT Trivector test for 20 right eyes are, respectively, plotted along the three axes, with labels p, d and t. Readers should allow their eyes to diverge to an imaginary point when observing each of the halves of the stereo pairs. This will result in a third part appearing in the middle between the two halves of the stereo-pair and this central part will have a three-dimensional appearance.

FIGURE 3: Stereo pairs of a colour space with 20 comets joining the Cambridge Colour Test - Trivector test and retest measurements. The origin represents 0 $u'v'$ units (or a chromatic Trivector [p d t]' of [0 0 0]') and the axis lengths are 0.01 $u'v'$ units.

test and retest trivectors for the eye concerned, and the lengths of the comets are a qualitative (and quantitative) measure of the similarity or equality of test and retest trivectors for that individual. Variations between test and retest results for different eyes then can be easily compared and, depending on the scale, shorter distances between two points show little variation in test and retest measurements. Points or comets that cluster within the same region will also indicate that measurements are similar in value.[27]

In Figure 3 there are 20 comets (one for each participant) and it can be seen that most of the test and retest measurements appear to be similar, with perhaps three or four exceptions, and particularly the comet and point indicated with the arrows where test and retest measurements are dissimilar for the participant concerned. This comet roughly parallels the tritan axis, suggesting that test and retest measurements for the tritan coordinates were more variable. Excluding this comet, all other comets are clustered closely together in the colour 3-space.

Figure 4 represents the test and retest stereo-pair scatter plots and 95% surfaces of constant probability density (approximately 95% of sample measurements are expected to be found within each ellipsoid).[27] Test and retest measurements are indicated by black or magenta points, respectively, and most measurements for the test and retest samples cluster about the means of the samples concerned. When viewing the scatter plots in three-dimensional space, both the test and retest measurements seem to cluster within the same region, indicating similarity, but both samples vary more with respect to the axis for the tritan coordinates, with the principal diameters of the ellipsoids almost paralleling this axis. The larger distribution ellipsoid for the retest measurements in magenta suggests greater variation of measurements compared to test measurements. However, an outlying value (the magenta dot above the magenta ellipsoid) is important

and its removal would result in the magenta ellipsoid becoming smaller and more similar to the black (test) one. In Table 2 some basic statistics for the two samples are provided.

Variances and *covariances* indicate the *variabilities* and *linear relationships*, respectively, between the test and retest trivectors (protan, deutan and tritan) in the 3-space. If no variation exists, then all six entries will be zero.[26] Overall variances were found to be slightly greater for the Trivector retest measurements (Tri-rt) but an outlying value was important. Also greater test and retest variation was found for the tritan coordinates (*t*). All covariances were close to zero; thus minimal or no linear relationships between paired variances exist, and variance for a specific vector along a confusion axis, for example, protan variance, is not linearly related to the variance along the confusion axes of either of the other two vectors (deutan or tritan).

Reliability statistics

Bland–Altman plots in Figure 5 and corresponding statistics (see Table 3) are used to investigate repeatability of the Trivector test by comparing test and retest measurements for the trivectors along three confusion lines for 20 right eyes and plotting their differences against their means.

The horizontal solid red line (Figure 5a) in each plot indicates the *mean difference* ($\bar{X}_d$) or estimated bias of the 20 paired measurements, with the distance from zero (no difference) providing an estimate of the average difference between the two methods. Theoretically, if both samples of measurements were identical and free from error, the mean difference would be zero and the further away the mean difference is from zero, the larger the magnitude of the bias; in this instance, as the test values were subtracted from the retest values, a positive mean difference would indicate that, on average, the retest values are slightly larger. The 95% LoA are indicated in Figure 5a using dashed lines, and the lines (solid or dashed) also include 95% confidence intervals (the shaded regions about the upper and lower LoA). These LoA were calculated using 1.96 times the standard deviation of the differences between the two samples (d_2 and d_1 where d_2 = retest and d_1 = test) under the assumption of normally distributed differences.

Mean differences $\bar{X}_d$ (see Figure 5 and Table 3) for the trivectors in the Bland–Altman plots are very close to zero, suggesting that on average, for these 20 right eyes, not much difference occurs between the test and retest results produced by the Trivector test. The means and differences are not correlated ($r = 0.42$, $p = 0.071$).

Note: The black points represent test measurements and the magenta points represent the corresponding retest measurements for 20 right eyes of adults aged 19–24 years. The axis lengths are 0.01 *u'v'* and the origin is at 0 *u'v'* units. Protan, deutan and tritan coordinate axes are indicated with labels *p*, *d* and *t*.

FIGURE 4: Stereo-pair scatter plots with 95% distribution ellipsoids for the Cambridge Colour Test - Trivector test and retest measurements.

TABLE 2: Means, variances, covariances and 95% distribution ellipsoids for the Trivector test and retest measurements.

Variable	Means	Variances			Covariances		
	(*p d t*)'	S_P	S_D	S_T	S_{PD}	S_{TP}	S_{TD}
Tri-t	(55 54 89)'	0.022	0.021	0.089	0.005	0.027	0.010
Tri-rt	(62 61 88)'	0.036	0.032	0.215	0.018	0.048	0.022

Note: Variances of protan (S_P), deutan (S_D) and tritan (S_T) vectors for 20 right eyes of adults aged 19–24 years are represented. S_{PD} represents the covariance of the protan and deutan vectors, S_{TP} represents the covariance of the tritan and protan vectors and S_{TD} represents the covariance of the tritan and deutan vectors. The units are *u'v'* × 10⁻⁴ for the means and (*u'v'*)² × 10⁻⁴ for variances and covariances.

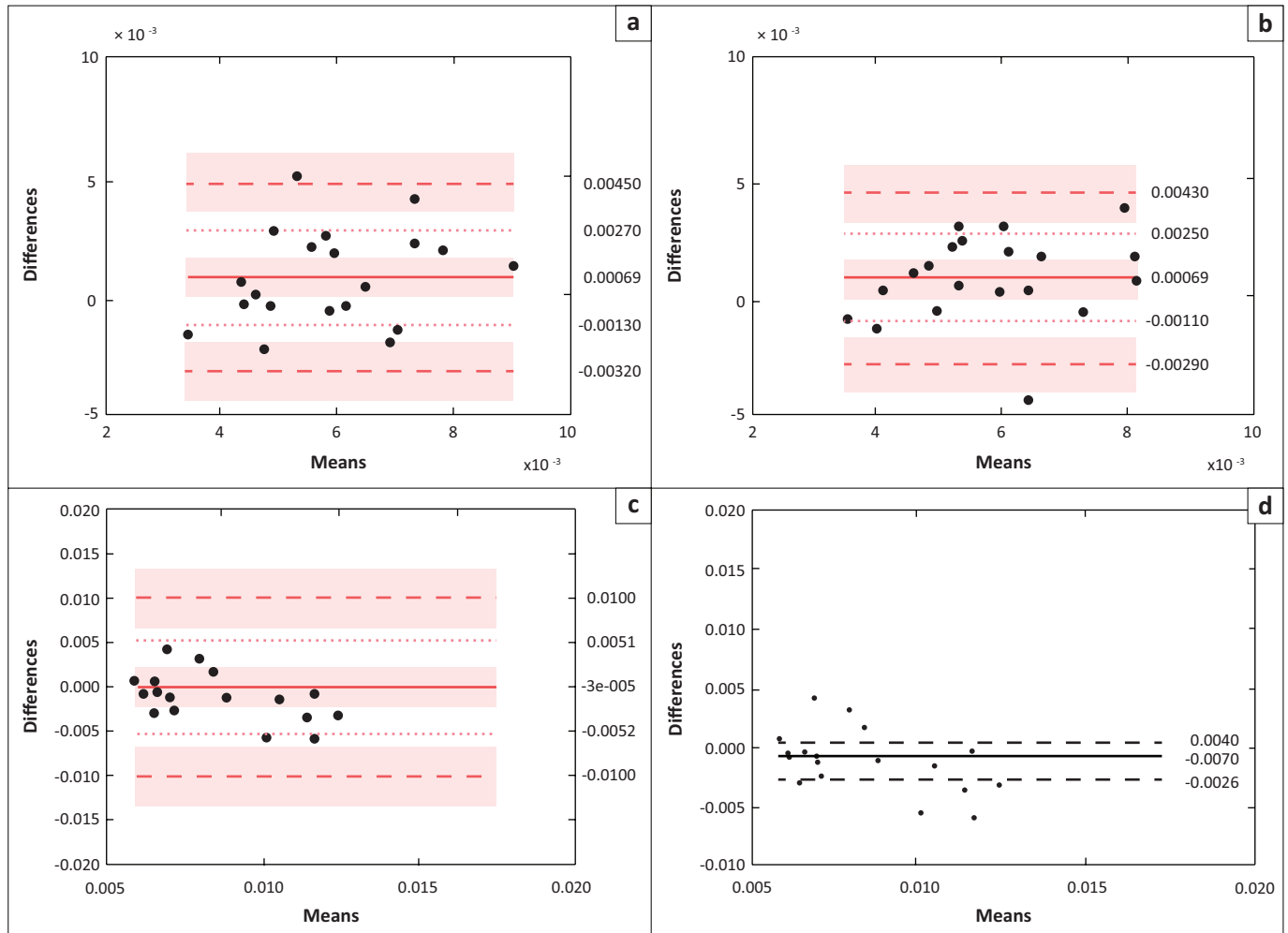Tri-t, Trivector test; Tri-rt, Trivector retest.

FIGURE 5: Bland–Altman plots of means versus differences and 95% limits of agreement (LoA) of test and retest for (a) protan, (b) deutan and (c) tritan CCT trivectors for 20 right eyes. Because of an outlier (see the point near the top of the plot), the plot for the tritan results has different scales. As a result, a non-parametric Bland–Altman plot of medians and 95% LoA for the non-normally distributed tritan differences was drawn for this sample and is shown in (d).

TABLE 3: Descriptive statistics for the Bland–Altman plots in Figure 5.

Variable	Protan	Deutan	Tritan	Tritan (non-normally distributed differences)
$\overline{X}$ (s.d.)	5.89 (1.38)	5.79 (1.35)	8.86 (2.93)	-
$\overline{X}_d$ (s.d. of differences)	0.69 (1.96)	0.69 (1.82)	-0.03 (5.15)	$\widetilde{M}_d$ -0.70
s.e. for $\overline{X}_d$	0.44	0.41	1.15	
(LLoA; ULoA)	(-3.24; 4.54)	(-2.87; 4.25)	(-10.12; 10.06)	(-2.60; 4.00)
s.e. for LoA	0.76	0.70	1.99	-
CR	0.01	0.01	0.01	-
CV$_{test}$	0.27	0.27	0.34	-
CV$_{retest}$	0.30	0.29	0.52	-

Note: The means ($\overline{X}$), mean differences ($\overline{X}_d$) median differences ($\widetilde{M}_d$), standard deviations (s.d. and s.d. of differences), standard errors, LLoA, ULoA (all the aforementioned quantities need to be multiplied by 10^{-3} $u'v'$ units), CR and CV for 20 right eyes of adults, aged 19–24 years, for the protan, deutan and tritan vectors for the CCT Trivector test.

$\overline{X}$, means; $\overline{X}_d$, mean differences; $\widetilde{M}_d$, median differences; s.d., standard deviations; s.e., standard errors; LoA, limits of agreement; LLoA, lower limits of agreement; ULoA, upper limits of agreement; CR, coefficients of repeatability; CV, coefficients of variation; CCT, Cambridge Colour Test.

When looking at the upper and lower LoA (Figure 5), it can be seen that the interval is slightly larger for the tritan vectors and this suggests the tritan vectors have slightly more variability in results yielded by the 20 subjects. Further calculations indicated that the tritan differences (between test and retest) were not normally distributed; therefore an additional Bland–Altman plot using medians (±95% LoA) was provided in Figure 5d.[28,29] The median difference (solid line) (Figure 5d and Table 3) was –0.0007 $u'v'$ units and the 95% confidence interval (dashed lines) was

(–0.0026; 0.0004) $u'v'$ units. The first number in the brackets is the lower limit and the second the upper limit; note the asymmetry in the interval relative to the solid line (median difference). The 95% confidence interval length for the LoA changed from 0.02 to 0.003 $u'v'$ in Figure 5c and d, respectively.

Because the International Organisation for Standardisation[30] recommends expressing repeatability in terms of standard deviations, the CR can be calculated as 1.96 (or 2) times the

standard deviation of the differences between the two samples (d_2 and d_1 where d_2 = retest and d_1 = test):[22]

$$CR = 1.96 \times \sqrt{\frac{\Sigma(d_2 - d_1)^2}{n-1}}.$$ [Eqn 2]

This coefficient can also be read from a corresponding Bland–Altman plot by subtracting the mean difference ($\bar{X}_d$) from the upper 95% LoA. If the differences between two samples are all zero, then CR = 0, and the greater the value for CR the less similar are the test and retest measurements.[22] Table 3 has CR = 0.01 for all samples, and this indicates excellent repeatability between test and retest samples.

The CV represents the ratio of the standard deviation (σ) of the population to the population mean (μ), and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other. The CV is defined as follows:[22]

$$CV = \frac{\sigma}{\mu}.$$ [Eqn 3]

Thus CV shows the extent of variability in relation to the mean of the sample or population. If the sample parameters (σ and μ) are unknown, then CV can be estimated as a ratio of the sample standard deviation divided by the sample mean.[22] If CV < 1, there is low variance between samples, for example test and retest samples, and high variance occurs when CV > 1.

Both the CR and CV in Table 3 suggest that the CCT Trivector test produces repeatable results with low variation occurring within the test and retest samples.

The stereo-pair in Figure 6 mimics an ordinary Bland–Altman plot but with a multivariate approach instead in order to assess repeatability.[20] The means of the test and retest trivectors for the 20 right eyes concerned are indicated by the cyan points enclosed mostly within the cyan 95% distribution ellipsoid. The differences of the retest and test measurements are similarly indicated by the magenta points mostly enclosed within the magenta 95% distribution ellipsoid. Corresponding results for each eye (that is, a 3 × 1 [vector] mean and ×1 difference of test and retest values) are linked to each other by a black comet or line segment and the centroid of the differences ellipsoid (in magenta) is indicative of the mean difference for the sample of 20 right eyes for the Trivector test. Similarly, the centroid of the means ellipsoid (in cyan) would provide an overall or global mean for the 20 right eyes for their test and retest trivectors.

Theoretically, if both Trivector measurements (test and retest for an eye) were identical and free from error, then only a single dot would be seen at the origin (a mean difference of zero). If the test and retest differences are small, then points will cluster close to the origin (zero difference) and the corresponding 95% differences ellipsoid will have a small volume. Similarly, the wider the spread of dots (cyan for differences and magenta for means) and the larger the corresponding 95% distribution
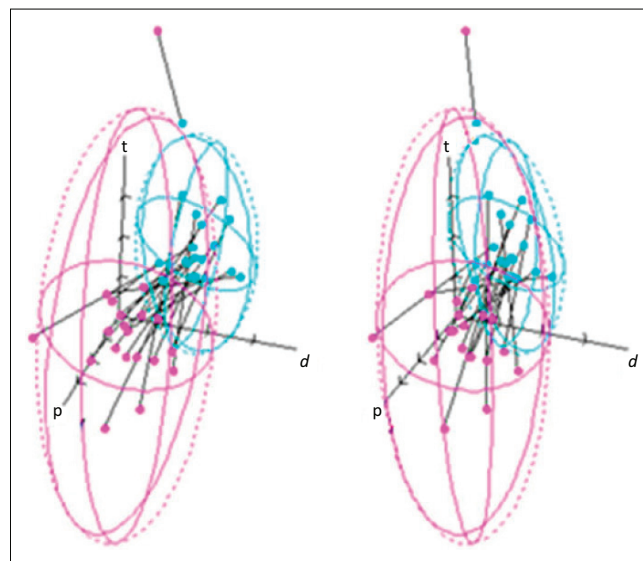

FIGURE 6: A stereo-pair scatter plot of means versus differences of test and retest measurements with 95% distribution ellipsoids for 20 adults aged 19–24 years. The cyan points (and ellipsoid) represent means of test and retest measurements, and the magenta points (and ellipsoid) represent the differences of test and retest measurements. The black comets link participant test and retest means to their corresponding test and retest differences. The origin is at 0 $u'v'$ units and the axis lengths are 0.01 $u'v'$ units.

TABLE 4: Variances (first three columns) and covariances of the 95% distribution ellipsoids for the means versus differences of test and retest measurements for the Cambridge Colour Test Trivector test for 20 right eyes of adults aged 19–24 years.

Variable	Variances			Covariances		
	S_P	S_D	S_T	S_{PD}	S_{TP}	S_{TD}
Trivector means	0.019	0.018	0.086	0.009	0.028	0.010
Trivector differences	0.039	0.033	0.265	0.010	0.035	0.022

Note: The units are ($u'v'$)$^2 \times 10^{-2}$ for both variances and covariances.

S_P, variance of protan vectors; S_D, variance of deutan vectors; S_T, variance of tritan vectors; S_{PD}, covariance of the protan and deutan vectors; S_{TP}, covariance of the tritan and protan vectors; S_{TD}, covariance of the tritan and deutan vectors.

ellipsoids, the greater the variation in the differences and/or means. The greater the differences and its ellipsoid, the weaker the repeatability between test and retest measurements (CCT trivectors in this case).[20] Although the ellipsoids in Figure 6 look relatively large, the axis lengths are only 0.01 $u'v'$ units. If, for example, axes of 1 $u'v'$ units were used instead, the ellipsoids would be seen to be very small. (Thus the scale applied needs to be taken into careful consideration when interpreting the meaning of the results.)

When analysing the stereo-pair scatter plots for the means of the samples versus their differences (see Figure 6 and Table 4), the Trivector tests for the 20 right eyes produce test and retest results that are consistent and repeatable. The differences in magenta can be seen as being more variable than the means in cyan. That is, the magenta points are more widely spread and their 95% distribution ellipsoid is larger than for the cyan points and their corresponding ellipsoid (an outlier was important here). However, the ellipsoids for both the differences and the means are more variable with respect to the tritan axis (and tritan coordinate vectors). Variances suggest higher variability with the tritan vectors. Covariances, although larger when the tritan vector is involved, are close to zero, indicating no linear relationships between the various coordinate vectors.

TABLE 5a: The Wilcoxon matched pairs hypothesis tests comparing test and retest measurements for 20 right eyes for the protan, deutan and tritan vectors of the Trivector test (see Figure 2).

Variable	T	Z	p
Pt and Prt	63.000	1.288	0.198
Dt and Drt	53.000	1.941	0.052
Tt and Trt	71.000	1.269	0.204

T, test statistic; Z, value for a 95% confidence interval; p, probability value; Pt, protan test; Prt, protan retest; Dt, deutan test; Drt, deutan retest; Tt, tritan test; Trt, tritan retest.

TABLE 5b: Hypothesis tests ($\alpha = 0.05$) of equality of means and of variances and covariances on Trivector test and retest vectors for 20 right eyes (see Figure 4).

Variable	Variances and covariances			Means		
	$\chi^2_{\alpha, 6}$	u	H_0	$F_{\alpha, 3, 36}$	w	H_0
Trivector	12.59	6.28	Accepted	2.84	1.04	Accepted

χ^2, sample statistic; u, critical value; F, sample statistic; w, critical value; H_0, null hypothesis.

Hypothesis tests

The Wilcoxon test is a non-parametric equivalent of the parametric paired t-test and is used to test for a difference in the mean (or median) of paired observations – before and after measurements with the same units and in Table 5, results represent the test and retest measurements for the Trivector tests for 20 right eyes. With this test, if the p-value is less than 0.05, we can reject the null hypothesis and the sample being compared, for example, test and retest measurements significantly differ.[31,32] From Table 5, we are confident in saying that test and retest measurements produced by the Trivector test are repeatable and reliable as the p-values of these tests are greater than the significance level of 0.05.

When hypothesis tests[33,34] for equality of means or equality of variances and covariances are applied to the test and retest Trivector data (Figure 4 and Table 5), the null hypothesis is accepted as equal variances, covariances and means are present, therefore implying that the Trivector test produces test and retest results for these 20 right eyes that are consistent and repeatable.

Discussion

Three samples (in Table 1) for the Trivector results were normally distributed and three were not normally distributed. A possible reason for the results for some of these samples (such as for the tritan vector being not normally distributed) could be the presence of outlying values in the test and/or retest measurements for the sample of 20 right eyes. Fatigue or lack of concentration perhaps could have influenced the tritan retest measurement as this participant (that is, the outlier) showed no tritan deficiency with the preliminary colour vision tests; therefore, a removal of this possible outlier from the sample would perhaps influence the distribution in terms of normality. The sample size ($n = 20$) was also relatively small and this might also have been a contributing factor.

The tritan vector medians, interquartile ranges and standard deviations, while very similar (for test and retest samples), were larger than those for the protan and deutan vector medians, suggesting that the participants were probably less

sensitive along the tritan confusion line and one or more outlying values were also important.

Using Bland–Altman plots and LoA (Figure 5 and Table 3), mean differences and the 95% LoA are small for the protan, deutan and tritan measurements, suggesting that differences and mean differences between test and retest measurements are small, and thus test and retest results for the Trivector test are in good agreement. Coefficients of repeatability and variation values and Bland–Altman LoA are more decisive for conclusions on repeatability and reliability,[22,32,35,36] and these measures indicate good agreement between test and retest for the Trivector test as values are close to zero.

Table 6 compares the means, standard deviations, medians and QDs (QD = ½ IQR) for results for the Trivector test from this article (in Johannesburg) to a study from Liverpool[15] for 40 subjects in the 20–29-year age group (mean ages were 22.6 ± 2.8 years) and combined data from a study in Brazil[16] for 75 subjects aged 18–30 years, that is, the São Paulo group of 45 subjects with mean ages of 22.5 ± 2.7 years and the Belem group of 30 subjects with mean ages of 21.8 ± 2.9 years.

From Table 6, results for the Trivector test show that means were similar for the groups in Liverpool[15] and Brazil,[16] although generally the means were lower in the Liverpool sample. Paramei[15] states that a factor that may result in some differences in statistical data could be the use of the CCT software, as the Belem group ran the CCT using a self-built system (IBM RISC 6000) and self-developed software, and the São Paulo and Liverpool groups employed the CCT commercial version (v2.0; VSG 5 card; CRS). In this study, Trivector means and standard deviations mostly were slightly larger (however, quantities in Table 6 are quite small and should be multiplied by 10^{-4}), and possible reasons for this could be influence of outlying values and the smaller sample size used in comparison to the other studies.

Medians are less influenced by outliers, and, if one compares the results in this study to the medians (in Liverpool) and means (in Liverpool and Brazil) of the other studies, the medians for the protan, deutan and tritan vectors are still larger, suggesting that outliers and/or sample size may have an influence on the inflation of results. In terms of standard deviations and QDs, results in this article are similar to the other studies for the deutan vectors but slightly larger for the protan vectors and particularly for the tritan vectors. Again, differences in sample sizes across the studies and the influence of outlying values are relevant issues. It can be seen, however, that across the studies, the tritan vector generally had larger colour values, suggesting that subjects are less sensitive along the tritan confusion line than along the protan or deutan confusion lines.

Test and retest measurements of the Trivector test were also plotted using stereo pairs and comets (Figure 6), and these figures are a relatively unique multivariate method where

TABLE 6: Comparisons of means and standard deviations for the Cambridge Colour Test Trivector test between studies (Brazil[16] and Liverpool[15]) and this study[20]; medians and quartile deviation are included for Johannesburg[20] and Liverpool[14].

Statistics†	Protan			Deutan			Tritan		
	Liverpool	Brazil	Johannesburg	Liverpool	Brazil	Johannesburg	Liverpool	Brazil	Johannesburg
Means	42.3	43.2	-	42.2	47.3	-	54.9	67.7	-
s.d.	9.5	11.4	-	11.8	15.0	-	15.4	19.6	-
Means (t)	-	-	55.4	-	-	54.5	-	-	88.8
s.d. (t)	-	-	14.7	-	-	14.4	-	-	29.9
Means (rt)	-	-	62.3	-	-	61.4	-	-	88.5
s.d. (rt)	-	-	18.9	-	-	18.0	-	-	46.4
Averaged (t and rt)	-	-	58.9 (16.8)	-	-	58.0 (16.2)	-	-	88.7 (38.2)
Medians	40.0	-	-	40.5	-	-	53.0	-	-
QD	6.5	-	-	10.0	-	-	8.0	-	-
Medians (t)	-	-	52.0	-	-	51.5	-	-	77.5
QD (t)	-	-	8.8	-	-	10.0	-	-	25.0
Medians (rt)	-	-	62.5	-	-	63.5	-	-	78.5
QD (rt)	-	-	13.8	-	-	12	-	-	16.3
Averaged (t and rt)	-	-	57.3 (11.3)	-	-	57.5 (11.0)	-	-	78.0 (20.7)

Note: Each statistic should be multiplied by 10^{-4} $u'v'$ units. For example, for the red (protan) vectors in Liverpool and Johannesburg, the means and standard deviations are 0.00423 ± 0.00095 $u'v'$ units and 0.00554 ± 0.00147 $u'v'$ units, respectively.
t, test; rt, retest; s.d., standard deviation; QD, quartile deviation.

test and retest trivariate data can be explored in a 3-space in terms of means and differences. This three-dimensional graphical approach has not previously been applied to Trivector results but has recently been used with other data such as contrast visual acuities.[37] However, comparisons of means and differences of test and retest measurements herein showed that the majority of the Trivector test and retest means cluster in front of and vary mainly in a direction almost parallel to the tritan axis, indicating greater variability for the tritan vectors. The differences between test and retest measurements mostly cluster near the origin (or zero difference), and the centroid of the ellipsoid (or mean difference, $\overline{X}_d$) is also close to the origin at zero, indicating good repeatability of data. Variances (see Table 3) confirm that the data for the tritan vectors for the sample of 20 right eyes are the most variable of the three vectors, and covariances are close to zero, indicating a lack of linear relationship between variances for any of the three colour vectors concerned. The volumes of the Trivector means and differences ellipsoids are also small, indicating consistency and good agreement; however, the differences can be seen as being more variable than the means.

Also, importantly, despite the relatively small sample size and the possibility of outliers, univariate and multivariate hypothesis tests found that the test and retest samples for all three vectors were not significantly different in terms of their means and standard deviations. Thus, Trivector results were reliable, and this is an important result in terms of supporting future clinical and research applications.

Conclusion

Through comprehensive analyses of Trivector results produced by the CCT on healthy human eyes with normal trichromatic vision, reliability and repeatability of the test and retest measurements were established. Although departure from normality and outlying values can have important influences, essentially, the CCT provides very similar test and retest results for the Trivector test.

The Trivector test thus can be used with confidence by researchers or clinicians, as it provides reliable results in repeated measures.

Acknowledgements

Competing interests

Authors' contributions

N.H. collected the data and wrote the manuscript as part of her master's dissertation, with supervision by A.R. Review of the initial draft, supervision and approval of the manuscript were performed by A.R.

References

1. Birch J. Diagnosis of defective colour vision. Oxford: Oxford University Press; 1996.
2. Dain S. Clinical colour vision tests. Clin Exp Optom. 2004;87:276–293. https://doi.org/10.1111/j.1444-0938.2004.tb05057.x
3. Cole BL. Assessment of inherited colour vision defects in clinical practice. Clin Exp Optom. 2007;90:157–175. https://doi.org/10.1111/j.1444-0938.2007.00135.x
4. Lakowski R. Theory of practice of colour vision testing: A review. Part 1. Br J Ind Med. 1969;26:73–189.
5. Lakowski R. Theory of practice of colour vision testing: A review. Part 2. Br J Ind Med. 1969;26:265–288.
6. Pease PL. Colour vision. In: Benjamin IM, Borish WJ, editors. Borish's clinical refraction. 2nd ed. Missouri: Butterworth-Heinemann, 2006; p. 289–355.
7. Paramei GV, Bimler DL. Color vision testing. In: Luo R, editor. Encyclopedia of color science and technology. New York: Springer, 2016; p. 407–413.
8. Paramei GV. Color perception and environmentally based impairments. In: Luo R, editor. Encyclopedia of color science and technology. New York: Springer, 2016; p. 343–348.
9. Mollon JD, Regan BC. Cambridge Colour Test. Handbook [homepage on the Internet]. Cambridge: Cambridge Research Systems Ltd., 2000 [cited 2017 May 17]. Avalable from: http://www.crsltd.com/tools-for-vision-science/measuring-visual-functions/cambridge-colour-test/
10. Hasrod N, Rubin A. Colour vision: A review of the Cambridge colour test and other colour testing methods. AVEH. 2015;74:1–7. http://doi.org/10.4102/aveh.v74i1.23
11. Mollon JD, Reffin JP. A computer controlled colour vision test that combines the principles of Chibret and Stilling. J Physiol. 1989;414:20.

12. Regan BC, Reffin JD, Mollon JD. Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. Vision Res. 1994;34:279–1299. https://doi.org/10.1016/0042-6989(94)90203-8

13. Shubert EF. Colorimetry. In: light-emitting diodes, 2nd ed., Cambridge: Cambridge University Press; 2006.

14. Paramei GV, Oakley B. Variation of color discrimination across the life span. J Opt Soc Am. 2014;29:A375–A384. https://doi.org/10.1364/JOSAA.31.00A375

15. Paramei GV. Color discrimination across four decades assessed by the Cambridge Colour Test. J Opt Soc Am. 2012;29:A290–A297. https://doi.org/10.1364/JOSAA.29.00A290

16. Ventura DF, Silveira LCL, Rodrigues AR, et al. Preliminary norms for the Cambridge Colour Test in normal and defective colour vision. In: Mollon JD, Pokorne J, Knoblauch K, editors. Normal and defective colour vision. Oxford: Oxford Press, 2003; p. 331–339.

17. Moura ALDA, Teixeira RAA, Oiwa NN, et al. Chromatic discrimination losses in multiple sclerosis patients with and without optic neuritis using the Cambridge Colour Test. Vis Neurosci. 2008;25:463–468. https://doi.org/10.1017/S0952523808080437

18. Silva MF, Faria P, Regateiro FS, et al. Independent patterns of damage within magno-, parvo- and koniocellular pathways in Parkinson's disease. Brain. 2005;128:2260–2271. https://doi.org/10.1093/brain/awh581

19. Mancuso K, Neitz M, Neitz J. An adaptation of the Cambridge Colour Test for use with animals. Vis Neurosci. 2006;23:695–701. https://doi.org/10.1017/S0952523806233364

20. Hasrod N. Colour discrimination and the Cambridge Colour Test: Reliability of discrimination Ellipses and Trivectors in colour space. MPhil dissertation, Johannesburg, South Africa: Department of Optometry, University of Johannesburg; 2015.

21. World Medical Association. Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA. 2013;310:2191–2194. https://doi.org/10.1001/jama.2013.281053

22. McAlinden C, Khadka J, Pesudovs K. Statistical methods conducting agreement (comparison of clinical tests) and precision (repeatability and reproducibility) studies in optometry and ophthalmology. Ophthalmic Physiol Opt. 2011;31:330–338. https://doi.org/10.1111/j.1475-1313.2011.00851.x

23. Razali NM, Wah YB. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. J Stat Model Anal. 2011;2:21–33.

24. Thode HC. Testing for normality. Boca Raton: CRC Press; 2002.

25. Tabachnick BG, Fidell LS. Using multivariate statistics, 3rd ed. New York: Harper Colins; 1996.

26. DeCarlo LT. On the meaning and use of kurtosis. Psychol Methods. 1997;2: 292–307. https://doi.org/10.1037/1082-989X.2.3.292

27. Malan DJ. Dioptric power data analysis: Computer implement of graphical methods with clinical examples. S Afr Optom. 1993;52:84–90.

28. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8:135–160. https://doi.org/10.1177/096228029900800204

29. Carkeet A, Goh YT. Confidence and coverage for Bland-Altman limits of agreement and their approximate confidence intervals. Stat Methods Med Res. 2016. https://doi.org/10.1177/0962280216665419

30. British Standards Institute (BSI) and International Organization for Standardization (ISO). Accuracy (trueness and precision) of measurement methods and results. 1994;5725:1–6.

31. Field A. Discovering statistics using SPSS, 3rd ed. California: Sage; 2011.

32. Bland M. An introduction to medical statistics, 3rd ed. New York: Oxford University Press; 2000; p. 201–281.

33. Harris WF. The mean and variance of samples of dioptric powers: The basic calculations. Clin Exp Optom. 1990a;73:89–92. https://doi.org/10.1111/j.1444-0938.1990.tb03110.x

34. Harris WF. Direct, vec and other squares, and sample variance-covariance of dioptric power. Ophthalm Physiol Opt. 1990b;10:363–372. https://doi.org/10.1111/j.1475-1313.1990.tb00883.x

35. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;327:307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

36. Bland JM, Altman DG. Statistical notes: Measurement error. Br Med J. 1996;313:744. https://doi.org/10.1136/bmj.313.7059.744

37. Sukha AY. Inter- and intra-subject variation of contrast visual acuities. DPhil thesis, Johannesburg, South Africa: Department of Optometry, Faculty of Health Sciences, University of Johannesburg; 2015.