Repeatability of Cone Contrast Color Vision Tests

Jeffery K. Hovis; Ali Almustanyir; Mackenzie Glaholt

INTRODUCTION: New computerized color vision tests are gaining popularity in the aviation community. These tests determine color vision status by measuring chromatic sensitivity and they can effectively classify color vision as normal vs. abnormal. However, little information is available regarding their repeatability. We evaluated the repeatability of two such tests: the Operational Based Visual Assessment Cone Contrast Test (OCCT) and the Rabin Cone Contrast Test (RCCT).

METHODS: A total of 56 subjects with normal color vision and 63 subjects with defective color vision completed both tests twice over 2 sessions. We determined the repeatability for a normal/abnormal result, between-eye differences in thresholds within a session, and between-session results for each eye.

RESULTS: Both tests had excellent repeatability for normal vs. abnormal color vision (i.e., using a cutoff score of 75 Rabin Color Contrast Sensitivity Units). The OCCT also had excellent repeatability for acceptable vs. unacceptable color discrimination (i.e., a cutoff score of 55), whereas the RCCT repeatability was lower. The RCCT's lower repeatability was because the between-eye and between-session Limits of Agreement for the color-defective subjects were approximately ±40 relative sensitivity units. In contrast, the Limits of Agreement for the OCCT ranged from ±10 to ±15.

DISCUSSION: These results reinforce the advantage of using a finer stimulus change when estimating cone thresholds in the clinical setting.

KEYWORDS: cone contrast, repeatability, color vision deficiency, color vision tests, limits of agreement.

Hovis JK, Almustanyir A, Glaholt M. *Repeatability of cone contrast color vision tests*. Aerosp Med Hum Perform. 2025; 96(4):287–295.

Over the last 20 yr, there has been a transition from printed color vision tests to computer-based tests within the aviation medical community. Two of the more common tests in aviation medicine are the Color Assessment and Diagnosis test (COL-AEGLIA Institute for Occupational Vision, Lelystad Airport, Netherlands) and the Rabin Cone Contrast test (RCCT) (Innova Systems, Burr Ridge, IL, United States). Both measure chromatic thresholds against a gray background. The major differences between the two tests are the number of hues presented, the presence of luminous noise, and the procedures for measuring threshold.[1,2] Both tests have excellent validity when screening for red-green color vision defects.[3–5] Konan recently introduced a newer cone contrast test version (ColorDx CCT-HD). The ColorDX CCT-HD is a commercial version of the Operational Based Visual Assessment (OBVA) Cone Contrast Test (OCCT) developed by the OBVA Laboratory with the U.S. Air Force Research Laboratory. The major differences between the OCCT and the RCCT are: 1) the OCCT has an expanded contrast range and finer changes in contrast, allowing for the precise measurement of thresholds for individuals with normal color vision (NCV); and 2) the OCCT uses a Landolt ring stimulus instead of the letter stimulus used by the RCCT.

At the time of this study, only the OCCT was available and, accordingly, this paper will compare the OCCT with the Innova version of the RCCT. Each test is very good for identifying individuals with congenital red-green color vision deficiencies (DCV)[6,7]; however, little information is available on the repeatability of each test. The coefficient of repeatability (COR) for RCCT cone sensitivity for NCV subjects is given as ±15 for the long wavelength sensitive cone (L-cone) and medium wavelength sensitive cone (M-cone). This value is based on the between-eye difference in sensitivity for NCV subjects.[4,8] Winterbottom et al.[9] recently argued that individuals with the maximum sensitivity in both sessions or eyes should be excluded from the COR analysis because the zero

From the School of Optometry and Vision Science, University of Waterloo, Waterloo, Ontario, Canada.

This manuscript was received for review in June 2024. It was accepted for publication in December 2024.

Address correspondence to: Jeffery K. Hovis, O.D., Ph.D., Associate Professor, School of Optometry, 200 University Ave. W, Waterloo, ON N2L 3G1, Canada; jhovis@uwaterloo.ca.

DOI: https://doi.org/10.3357/AMHP.6538.2025

difference is due to a limitation of the test's range and not a reflection of their accuracy. If these individuals are excluded, the COR increases to ±30.[9] This last point raises the issue of whether the COR, based on the standard deviation of the differences for all patients, is the proper metric, given that the between-eye or between-day differences for individuals with NCV and perhaps individuals with red-green DCV are skewed due to the ceiling effect imposed by the limited contrast range of the RCCT. Because the OCCT uses a more extensive range of contrasts than the RCCT, it could be less prone to ceiling and floor effects, and different COR values might be expected. Differences in the psychophysical thresholding procedure between the OCCT and RCCT could also influence the repeatability of the cone sensitivity values and, consequently, the repeatability of pass/fail outcomes.

There are several reasons for quantifying the between-eye repeatability within and across sessions. These include determining whether the difference between eyes could be due to an ocular disease or disorder. If the between-eye difference within a session exceeds the COR, then there could be an underlying ocular disease present in the worse eye. The COR helps determine whether an acquired DCV is progressing in one or both eyes and how to manage an applicant who fails the test with one eye but passes with the other eye. If the difference is within the COR, one might be more inclined to qualify the candidate or at least retest. Also, because the pilot applicant who fails the test is often highly motivated, they frequently request another try or shop around for another test site if they fail the test. Knowing the between-session COR for the cone sensitivities and the between-session repeatability of the test on a pass/fail basis helps manage cases that failed and are requesting or seeking another attempt.

In addition to examining the repeatability of the threshold values, it is also essential to look at the repeatability of the tests on a pass/fail basis, especially since there are two proposed cutoff scores for the RCCT: 75 and 55. The 75 cutoff value is intended to identify all individuals with a congenital red-green color vision defect, while the 55 score is applied within military aviation to determine whether a person has adequate color vision to perform pilot color vision demands based on simulated cockpit displays and signal lights.[3,10] There is little published repeatability data on the OCCT and RCCT using these cutoff scores, and this aspect will also be examined in the present study; however, we will only be examining the repeatability of the L and M cone contrast sensitivities since these are the primary values of interest for aircrew.

METHODS

Subjects

The study received ethics clearance through the Office of Research Ethics at the University of Waterloo (ORE 20,996) and the Defense Research and Development Canada Human Research Ethics Committee (Protocol 2014-044). The results presented in this paper were from a larger study examining the performance of various color vision tests. The data analyzed were from 56 individuals (45.5% men) with NCV and 63 individuals (89.7% men) with DCV who participated in 2 sessions on different days. They were civilians recruited through posters, social media, and newsletter advertisements. The average age for the NCV was 26.9 (SD ± 9.7) and 28.0 yr (SD ± 10.7) for the DCV subjects. Color vision was classified according to the Rayleigh color match using the HMC Oculus anomaloscope (Oculus Optikgeräte GmbH, Wetzlar, Germany) in the neutral adaptation mode. The DCV group contained 63 individuals, of which there were 7 deuteranopes, 29 deuteranomalous, 19 protanopes, and 8 protanomalous. One subject was classified as "pigmentfarbeamblopie" based on his normal settings on the anomaloscope and failing all the other color vision tests used in this study.[11] He was placed in the deuteranomalous group because his results on the other tests were typical of a deutan defect.

Tinted contact lenses or spectacles were not allowed. Subjects were screened for ocular diseases using a short questionnaire. Visual acuities had to be at least 6/6 in the better eye and 6/9 in the other eye at 6 m, 0.8 M in the better eye, and 1.0 M in the other eye at 40 cm, either corrected or uncorrected. The acuity criteria were based on the Royal Canadian Air Force requirements, but the civilian Category 1 license criteria (i.e., commercial) are similar, with a minimum acuity of 6/9 in each eye and 6/6 binocularly.

Equipment

The RCCT was the commercial version 16.02.0 supplied by Innova Systems (Burr Ridge, IL). The stimuli were Sloan letters that subtend 1.6° at the 60-cm viewing distance (i.e., 20/385). The letters are presented individually in the center of the computer screen for up to 5 s, and the subject indicates which letter is presented by using a mouse to select the letter from the key displayed on the monitor. The subject could respond any time within the 5-s window and a new trial would begin after the response had been entered. The RCCT test was displayed on a 27.9-cm Acer laptop using the Windows 7 operating system. This program version required that the monitor be calibrated weekly using a Spyder colorimeter (Express ver. 4.5.4; Datacolor, Lawrenceville, NJ, United States). The white reference had a correlated color temperature of 6500 K. The luminance of the gray background was 19 $cd \cdot m^{-2}$. The test can present up to five contrast levels and up to five letters at each contrast. Cone contrast sensitivity was determined using a proprietary staircase procedure to give sensitivity scores similar to previous versions. Each eye was tested separately, with the right eye tested first.

The OCCT stimuli are Landolt rings, which subtend 1.4° with a gap of 0.3° at a 1-m viewing distance (i.e., 20/335). The subject's task was to identify the location of the gap using the keyboard arrows. Thresholds for each cone mechanism were determined using a four-alternative forced choice with the Psi adaptive method.[12] The psychometric function's slope was fixed at 2.6 for the L- and M-cone thresholds. This value was based on data from the OBVA Laboratory. The trials began with a practice session where suprathreshold stimuli were presented.

There were eight presentations for each cone stimulus. Next, the contrast threshold for each cone was calculated after 20 additional presentations. The right eye was tested first. The three different cone stimuli were presented randomly. The OCCT program (ver. 1.1.0) was run on a desktop (Lenovo Intel Core i5) with the Windows 7 Professional operating system. The stimulus was presented for up to 3 s on an NEC monitor (model 232 W-BK). The subject could respond anytime within the 3-s window and then a new trial would begin after the response was entered. The luminance of the gray background was 69 cd · m^{-2}. The monitor was calibrated using an X-Rite (ver. EODIS3 i1; Grand Rapids, MI, United States) Display Pro colorimeter every 30 d. Besides the difference in the computer hardware between the prototype and the Konan commercial version, the number of presentations for determining the threshold in the prototype was 20, whereas the commercial version has 30 presentations. Also, the order of the cone stimuli was randomly selected in the prototype rather than the L, M, and S-cone sequence in the commercial version.

The order of the color vision tests was determined using a random block design for the first session. The reverse order was used at the second session, held 10–15 d later. The monitors' warm-up time before collecting data or calibration was 15 min. for the LCD/LED displays.

Statistical Analysis

First, we examined the repeatability of pass/fail outcomes based on the 75 and 55 sensitivity criteria. Because each eye was tested, there could have been discrepancies between eyes regarding passing or failing. If this occurred in a session, the overall result was considered a failure since a discrepancy between eyes could indicate a borderline result, and additional testing is likely required. The kappa coefficient (κ) of agreement was used to calculate the pass-fail agreement between sessions for the two cutoff scores. The κ value can range from −1.0 for total disagreement, zero for chance agreement, and 1.0 for perfect agreement. AgreeStat (AgreeStat Analytics, Gaithersburg, MD, United States) was used to calculate the agreement between sessions. All other statistical analyses were completed using SPSS (ver. 29, IBM Corp, Armonk, NY, United States).

We had to scale the L- and M-cone contrast values between the RCCT and OCCT tests to compare the tests at these two criteria. The RCCT results are considered relative contrast sensitivity, with a 100 corresponding to the minimum contrast presented. The OCCT, however, uses an absolute scale to measure cone sensitivity, so a common scale was required. Because the RCCT COR is expressed in relative units, we converted the OCCT into relative cone sensitivity values. This conversion is not straightforward based on previous studies because there are two different equations for relating the commercial version of the OCCT and RCCT sensitivities, depending on which version of the RCCT is used, and the newest version of the RCCT has a slightly different minimum contrast than the version used in this study,[9,13] which complicates this conversion. Because the 75 and 55 cutoff scores were based on a similar version of the RCCT as used in this study, we derived a conversion that set the contrasts of the OCCT at sensitivities of 75 and 55 to be equal to the RCCT contrasts at 75 and 55. These contrasts were −1.60 log units and −1.28 log units for a sensitivity of 55. These values were derived based on our version of the RCCT, the manufacturer's specifications, and communications with the manufacturer and are slightly lower than published for the newer RCCT, which is likely due to the difference in the minimum contrast.[13]

The equation for converting the OCCT cone thresholds to the relative sensitivity scale is:

$$\text{Relative Cone Contrast Sensitivity} = 62.5\left(-\log\left(\text{cone contrast}_{\text{OCCT}}\right)\right) - 25 \quad \text{Eq. 1}$$

The differences between our contrasts from the newer version of the RCCT and the commercial version of the OCCT are partially due to a lower maximum contrast used in our version of the RCCT.

Next, we considered the repeatability of contrast sensitivity scores using a variation of the Limits of Agreement (LOA) method that considers the lack of normality in the sensitivity differences. For normally distributed data, the LOA is defined as the mean difference ±1.96 times the standard deviation of the difference. By comparison, the COR, which is applied in previous work,[4,8] is defined as ±1.96 times the standard deviation of the difference. The difference between the LOA and COR is that the LOA uses the actual mean difference in the calculation, whereas the COR assumes that the mean difference is zero. That is, if the mean difference is zero, then the LOA and COR will be equal. Regardless of whether the mean difference is zero, 95% of the differences fall within ±LOA or ±COR.

Winterbottom et al.[9] found that the distribution of the RCCT differences is skewed and leptokurtic because of the ceiling effect at the lowest contrast (highest sensitivity) stimulus presented in the test. The Shapiro-Wilk statistical test for normality confirmed that was also the case for our RCCT results. **Table I** shows that all the NCV RCCT sensitivity differences varied significantly from a normal distribution. Similarly, most of the protan subgroup differences varied considerably from a normal distribution, as did three out of four of the L-cone differences for the deutan subgroup. Interestingly, all the M-cone differences for the deutan subjects met the normality assumptions. The median differences will be reported because most difference distributions differed significantly from a normal distribution. Accordingly, we report 2.5 percentile and 97.5 percentile values as the nonparametric equivalents of the LOA for the NCV and the DCV deutan subgroup. For the DCV protan subgroup, 2.5 percentile and 95 percentile LOAs will be reported because the confidence intervals are calculated using a bootstrapping procedure in SPSS,[14] and there was insufficient data to calculate the 97.5 percentile confidence interval.

As can be seen in **Table II**, more of the OCCT between-eye and between-session differences in sensitivity met the normality assumption, but some did not: at least one of the L-cone differences did not meet the normality assumption in each of the three subject groups. In most cases, the M-cone sensitivity

Table I. Shapiro-Wilk Normality Test for Contrast Sensitivity Differences Between Eyes and Between Sessions for the RCCT.

COMPARISON	SHAPIRO-WILK W	DEGREES OF FREEDOM	*P*-VALUE
Color-Normals			
L-cone RE-LE 1st session	0.897	56	<0.001
L-cone RE-LE 2nd session	0.888	56	<0.001
L-cone RE 1st–2nd session	0.721	56	<0.001
L-cone LE 1st–2nd session	0.812	56	<0.001
M-cone RE-LE 1st session	0.649	56	<0.001
M-cone RE-LE 2nd session	0.648	56	<0.001
M-cone RE 1st–2nd session	0.760	56	<0.001
M-cone LE 1st–2nd session	0.592	56	<0.001
Deutans			
L-cone RE-LE 1st session	0.861	36	<0.001
L-cone RE-LE 2nd session	0.951	36	0.115
L-cone RE 1st–2nd session	0.912	36	0.007
L-cone LE 1st–2nd session	0.913	36	0.008
M-cone RE-LE 1st session	0.963	36	0.268
M-cone RE-LE 2nd session	0.943	36	0.062
M-cone RE 1st–2nd session	0.959	36	0.205
M-cone LE 1st–2nd session	0.952	36	0.122
Protans			
L-cone RE-LE 1st session	0.926	27	0.056
L-cone RE-LE 2nd session	0.738	27	<0.001
L-cone RE 1st–2nd session	0.909	27	0.022
L-cone LE 1st–2nd session	0.949	27	0.208
M-cone RE-LE 1st session	0.866	27	0.002
M-cone RE-LE 2nd session	0.834	27	<0.001
M-cone RE 1st–2nd session	0.875	27	0.004
M-cone LE 1st–2nd session	0.888	27	0.007

RE = right eye; LE = left eye; RCCT: Rabin Cone Contrast Test.

Table II. Shapiro-Wilk Normality Test for Contrast Sensitivity Differences Between Eyes and Between Sessions for the OCCT.

COMPARISON	SHAPIRO-WILK W	DEGREES OF FREEDOM	*P*-VALUE
Color-Normals			
L-cone RE-LE 1st session	0.986	56	0.749
L-cone RE-LE 2nd session	0.910	56	<0.001
L-cone RE 1st–2nd session	0.965	56	0.110
L-cone LE 1st–2nd session	0.955	56	0.037
M-cone RE-LE 1st session	0.967	56	0.132
M-cone RE-LE 2nd session	0.980	56	0.480
M-cone RE 1st–2nd session	0.991	56	0.953
M-cone LE 1st–2nd session	0.972	56	0.226
Deutans			
L-cone RE-LE 1st session	0.987	36	0.946
L-cone RE-LE 2nd session	0.984	36	0.867
L-cone RE 1st–2nd session	0.910	36	0.006
L-cone LE 1st–2nd session	0.962	36	0.242
M-cone RE-LE 1st session	0.967	36	0.351
M-cone RE-LE 2nd session	0.975	36	0.589
M-cone RE 1st–2nd session	0.976	36	0.618
M-cone LE 1st–2nd session	0.989	36	0.975
Protans			
L-cone RE-LE 1st session	0.812	27	<0.001
L-cone RE-LE 2nd session	0.925	27	0.051
L-cone RE 1st–2nd session	0.850	27	0.001
L-cone LE 1st–2nd session	0.833	27	<0.001
M-cone RE-LE 1st session	0.936	27	0.096
M-cone RE-LE 2nd session	0.958	27	0.340
M-cone RE 1st–2nd session	0.962	27	0.412
M-cone LE 1st–2nd session	0.897	27	0.012

RE = right eye; LE = left eye; OCCT: Operational Based Visual Assessment Cone Contrast Test.

differences met the normality assumption, except for the protan subgroup between-session differences for the left eye. Because many of the OCCT L-cone sensitivity differences did not meet the normality assumption, and to be consistent with the RCCT analysis, the corresponding nonparametric values were also determined for the OCCT.

RESULTS

Fig. 1 shows the between-session pass/fail repeatability for the cone sensitivity pass value of ≥75 for both the RCCT and OCCT. The κ-value for agreement for the RCCT test was 0.97 and the value for the OCCT was marginally lower at 0.95. Nevertheless, the confidence intervals overlapped, so the differences were not statistically significant, indicating that each test's between-session pass/fail outcome was excellent after correcting for chance. Although none of the NCV subjects had a pass/fail discrepancy between eyes at either session on the RCCT, five DCV subjects (7.9%) had a between-eye pass/fail discrepancy at the first session. The between-eye pass/fail results for these individuals at the second session were: three failed with each eye individually, one obtained passing scores in each eye, and one had the same between-eye pass/fail discrepancy. Four additional DCV subjects had between-eyes pass/fail discrepancies at the second session, and all four had failures for each eye at the first session. In addition, approximately 10% of the DCV subjects who failed in each eye for one cone stimulus at each session had between-eye pass/fail discrepancies for the other cone stimulus.

There were no between-eye pass/fail discrepancies for the NCV group on the OCCT at either session. Only one deuteranomalous DCV subject had a between-eye pass/fail discrepancy. This person failed each eye individually at the first session but only failed for the left eye during the second session. The RCCT pass/fail outcome for this individual was identical. Similar to the RCCT, approximately 15% of the DCV subjects who failed in each eye for one OCCT cone stimulus had between-eye pass/fail discrepancies for the other cone stimulus.

RCCT TEST	1st Session	
2nd Session	Pass	Fail
Pass	57 (NCV=55)	2 (NCV=1)
Fail	0 (NCV=0)	60 (NCV=0)
κ =0.97 (95% CI 0.92 to 1.00)		

OCCT TEST	1st Session	
2nd Session	Pass	Fail
Pass	56 (NCV=54)	2 (NCV=1)
Fail	1 (NCV=1)	60 (NCV=0)
κ =0.95 (95% CI 0.89 to 1.00)		

Fig. 1. Pass/fail repeatability using a cutoff score of ≥75 for the RCCT (top) and the OCCT (bottom) along with the kappa coefficient for agreement between sessions. CI: confidence interval; OCCT: Operational Based Visual Assessment Cone Contrast Test; RCCT: Rabin Cone Contrast Test.

RCCT TEST	1st Session	
2nd Session	Pass	Fail
Pass	62 (NCV=55)	2 (NCV=0)
Fail	5 (NCV=0)	50 (NCV=0)
κ =0.88 (95% CI 0.80 to 0.97)		

OCCT TEST	1st Session	
2nd Session	Pass	Fail
Pass	59 (NCV=56)	1 (NCV=0)
Fail	1 (NCV=0)	58 (NCV=0)
κ =0.97 (95% CI 0.92 to 1.00)		

Fig. 2. Pass/fail repeatability using a cutoff score of ≥55 for the RCCT (top) and the OCCT (below) along with the kappa coefficient for agreement between sessions. CI: confidence interval; OCCT: Operational Based Visual Assessment Cone Contrast Test; RCCT: Rabin Cone Contrast Test.

Fig. 2 shows the repeatability results for the cutoff score of ≥55. The κ coefficient of agreement between sessions is lower for the RCCT relative to the OCCT using the same contrast sensitivity cutoff and lower relative to the RCCT using the 75 cutoff score. Because the confidence intervals of the higher κ coefficients do not contain the RCCT κ coefficient of 0.88, the κ coefficient for RCCT repeatability using a score of 55 is significantly lower than the other κ-values for repeatability.

Although there were no between-eye pass/fail discrepancies for the NCV group on either test, the number increased for the DCV group for both tests. For the RCCT, the number of between-eye discrepancies increased to 14 DCVs. Nine subjects failed for just one eye in the first session: four of these individuals failed for one eye only at the second session (not necessarily the same eye as the first session), four failed with each eye individually at the second session, and one passed with each eye separately at the second session. Of the remaining five subjects, three failed each eye individually in the first session but failed with only one eye in the second session. Two individuals passed with each eye at the first session and failed for one eye at the second session.

The number of between-eye discrepancies on the OCCT increased to three DCVs. A subject failed with just one eye in the first session and failed on the same eye only in the second session. For the remaining two subjects, one failed on each eye individually at the first session and failed for only one eye at the second session, and one passed with each eye at the first session and failed for one eye at the second session.

It is possible that a difference in acuity could contribute to the between-eye discrepancies. Of the 14 subjects with between-eye discrepancies, 7 also had between-eye differences in their intermediate or near acuities. Three of these subjects had RCCT contrast sensitivity differences that were consistent with the acuity difference (i.e., lower acuity, lower contrast sensitivity), three had inconsistent differences (i.e., lower acuity, higher contrast sensitivity), and one had higher sensitivity in different eyes in the two trials. The relationship between the acuity and cone contrast sensitivity for the OCCT is similarly uncertain. Of the three subjects, two of these individuals had equal acuity in each eye at intermediate and near distances, and the third had better acuity in the eye with the lower cone contrast sensitivity.

Fig. 3 shows the median between-eye and between-session differences in cone contrast sensitivity for the RCCT, and **Fig. 4**

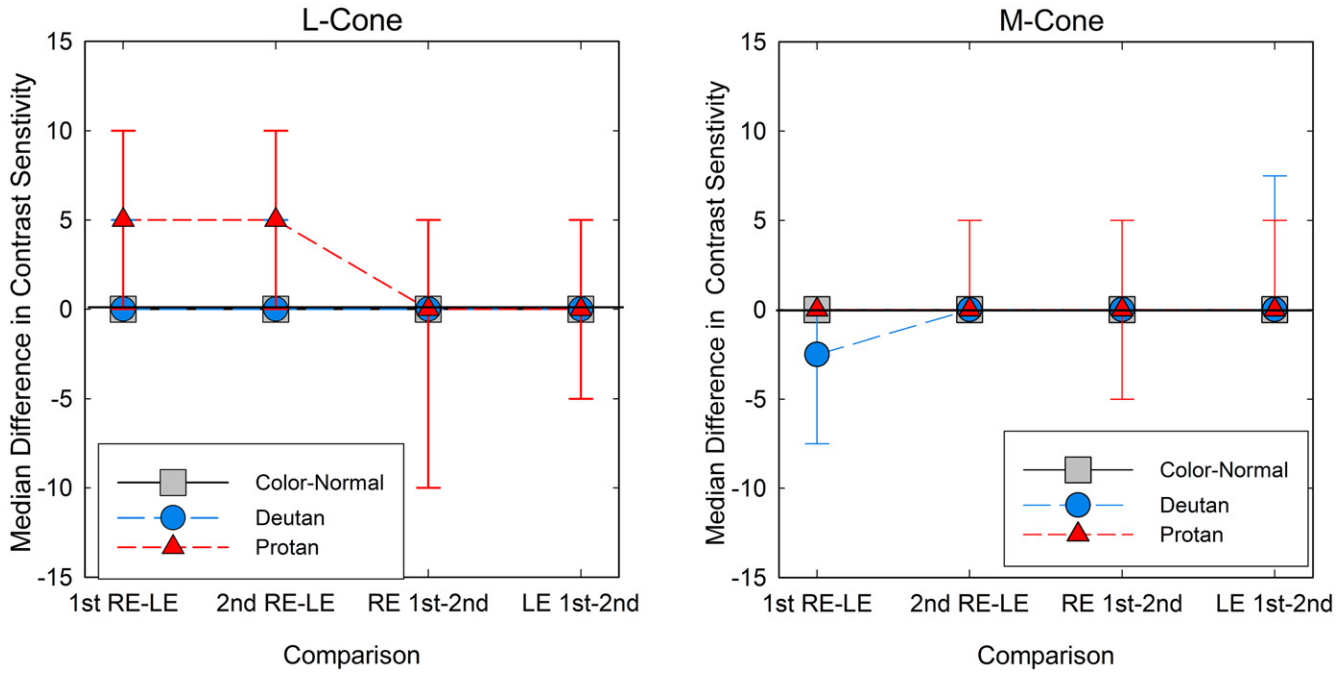

Fig. 3. RCCT median differences for the various comparisons on the x-axes. At the first session, 1st RE-LE is the between-eye comparison, 2nd RE-LE is the between-eye comparison at the second session, RE 1st–2nd is the comparison between first and second session for the right eye, and LE 1st–2nd is the comparison between the first and second session for the left eye. Error bars represent the 95% confidence interval for the median difference. Missing error bars indicate that the interval was equal to the median. RCCT: Rabin Cone Contrast Test.

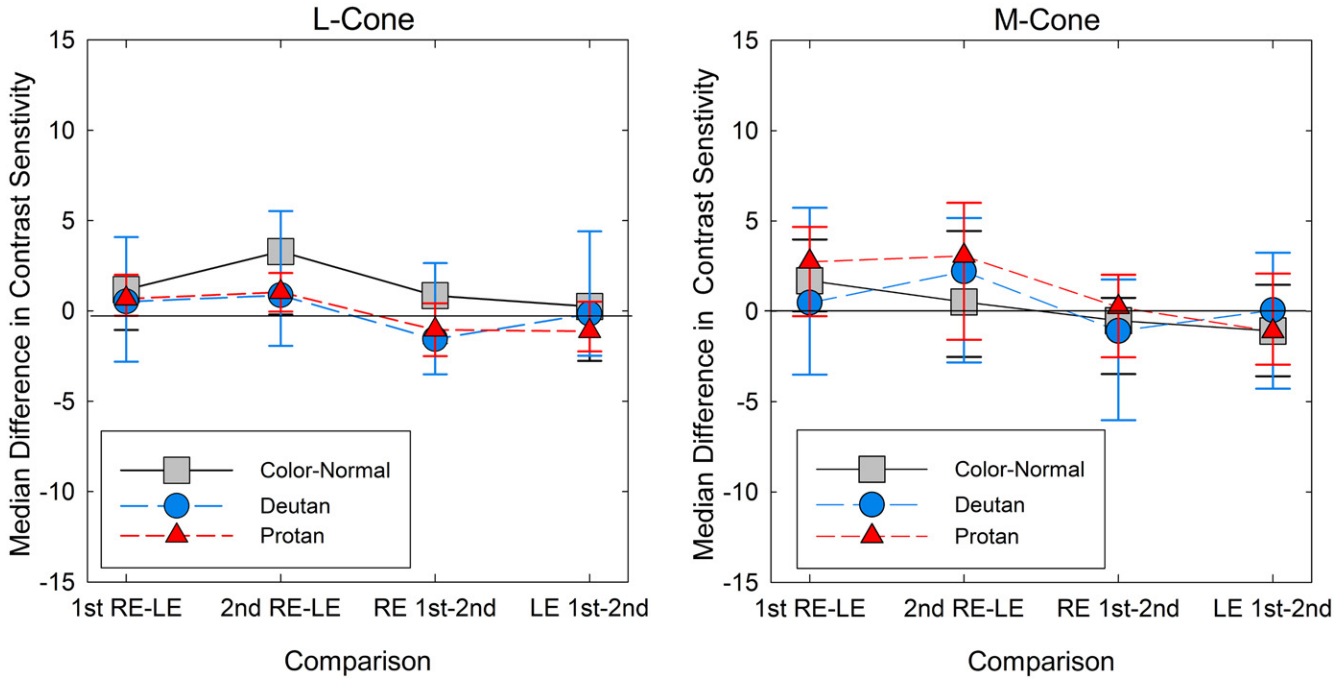

Fig. 4. OCCT median differences for the various comparisons on the x-axes. 1st RE-LE is the between-eye comparison at the first session, 2nd RE-LE is the between-eye comparison at the second session, RE 1st–2nd is the comparison between first and second session for the right eye, and LE 1st–2nd is the comparison between the first and second session for the left eye. Error bars represent the 95% confidence interval for the median difference. OCCT: Operational Based Visual Assessment Cone Contrast Test.

shows the corresponding values for the OCCT. For both tests, all the confidence intervals contain zero, so there is no statistically significant systematic difference between eyes or between sessions. In several RCCT comparisons for the NCV group, the confidence interval was zero. This result was because the bootstrap used a larger number of samples (1000) and the ceiling effect present in the test. For example, 73% of the NCV had a between-eye difference in M-cone sensitivity of zero at the first session because the sensitivity in each eye was 100. With $N = 1000$ samples used in the confidence interval calculation, both the upper and lower values of the interval approached zero.

Fig. 5 shows the LOA for the RCCT expressed as percentile values. In all cases, the median value was used to calculate the LOA. The COR of ±15 from previous studies is included for reference.[4,8] For the NCV group, the LOAs based on percentile values agree reasonably well with the ±15 for the L-cone differences. Except for the between-session right eye comparison, the 95% confidence intervals for the other comparisons contain ±15. The M-cone LOAs are generally smaller and close to 8; however, because the score is rounded to the nearest 5, the LOA of agreement should be increased to 10 for the M-cone.[15] On the other hand, the LOAs for the DCV group are larger, especially for the cone stimulus that corresponds to their defect.

Fig. 6 shows the LOAs for the OCCT differences. Similar to the RCCT, the NCV LOAs generally agree with the COR of ±15, but unlike the RCCT, the ±15 can also be applied to the deutans. The protan lower LOAs are generally smaller than the NCV and deutans for both the L-cone and M-cone sensitivities. The protan upper LOAs for the L-cone are also usually smaller, and part of this result is due to using the 95 percentile for the upper LOA instead of the 97.5 percentile to allow for an estimate of the precision of all LOAs. However, if the 95 percentile value is used for the deutans, the deutan upper LOAs remain larger for the between-eye differences at the second session and the right eye between-session differences. This is also the case for the NCV for the between-eye difference at the second session and the left-eye between-session difference. This asymmetry in the protan LOAs is from the positive skew in the difference distributions.

DISCUSSION

The repeatability of the two cone threshold tests is extremely good for a pass/fail cutoff score of ≥75. Both tests agree that a pass or fail result will be repeated on 97% of the individuals examined with each test. This result is expected since both tests have high validity in screening for red-green color vision defects.[4,8] The OCCT repeatability using the 55 cutoff score is also excellent, but the RCCT repeatability drops significantly, with 88% of the pass/fail results agreeing on both sessions. The likely reason for lower repeatability is the larger LOAs for the DCV group, which will be discussed next.

Although the RCCT repeatability was specified using nonparametric statistics, the between-eye LOAs for the L-cone sensitivity include the previous COR of ±15 for the NCV group. Our results show that the between-session LOAs for NCV L-cone sensitivity were not statistically different from

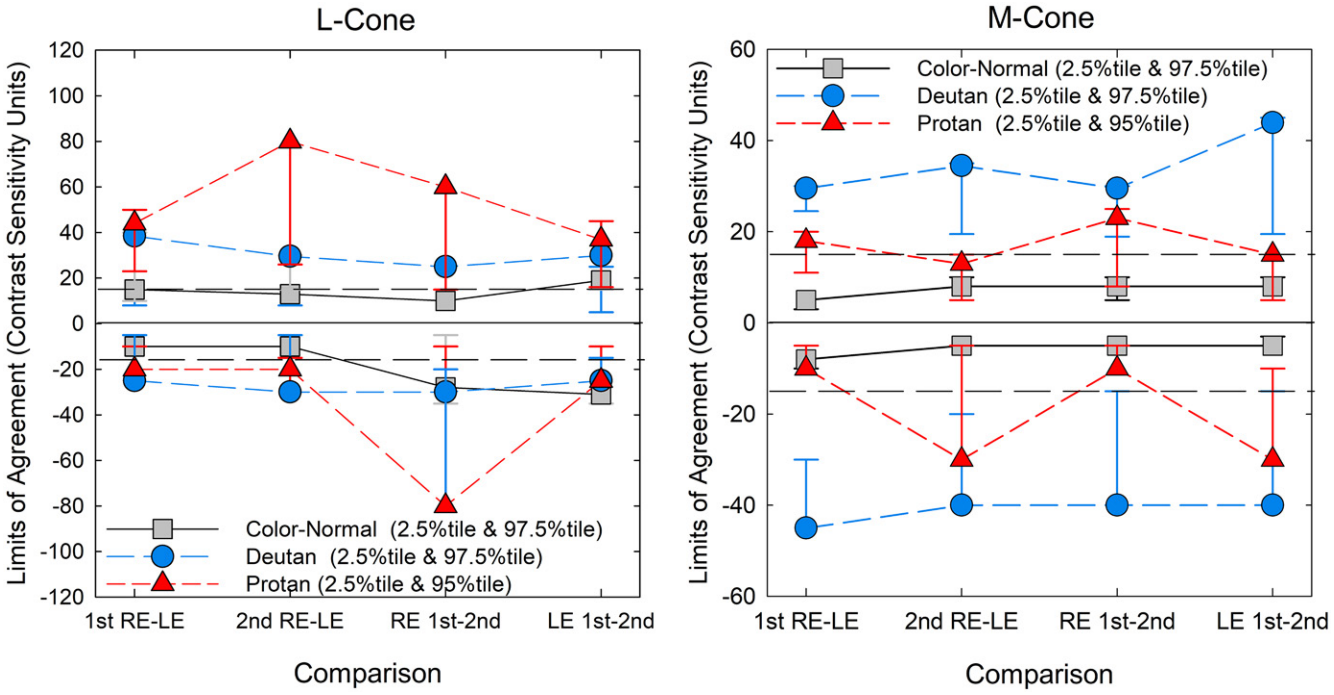

Fig. 5. RCCT limits of agreement for the various comparisons on the x-axis. 1st RE-LE is the between-eye comparison at the first session, 2nd RE-LE is the between-eye comparison at the second session, RE 1st–2nd is the comparison between first and second session for the right eye, and LE 1st–2nd is the comparison between the first and second session for the left eye. The dashed lines represent the COR of ±15 from previous studies.[4,8] Error bars represent the 95% confidence interval for the limits of agreement. Missing error bars indicate that the interval was zero. RCCT: Rabin Cone Contrast Test; COR: coefficient of repeatability.

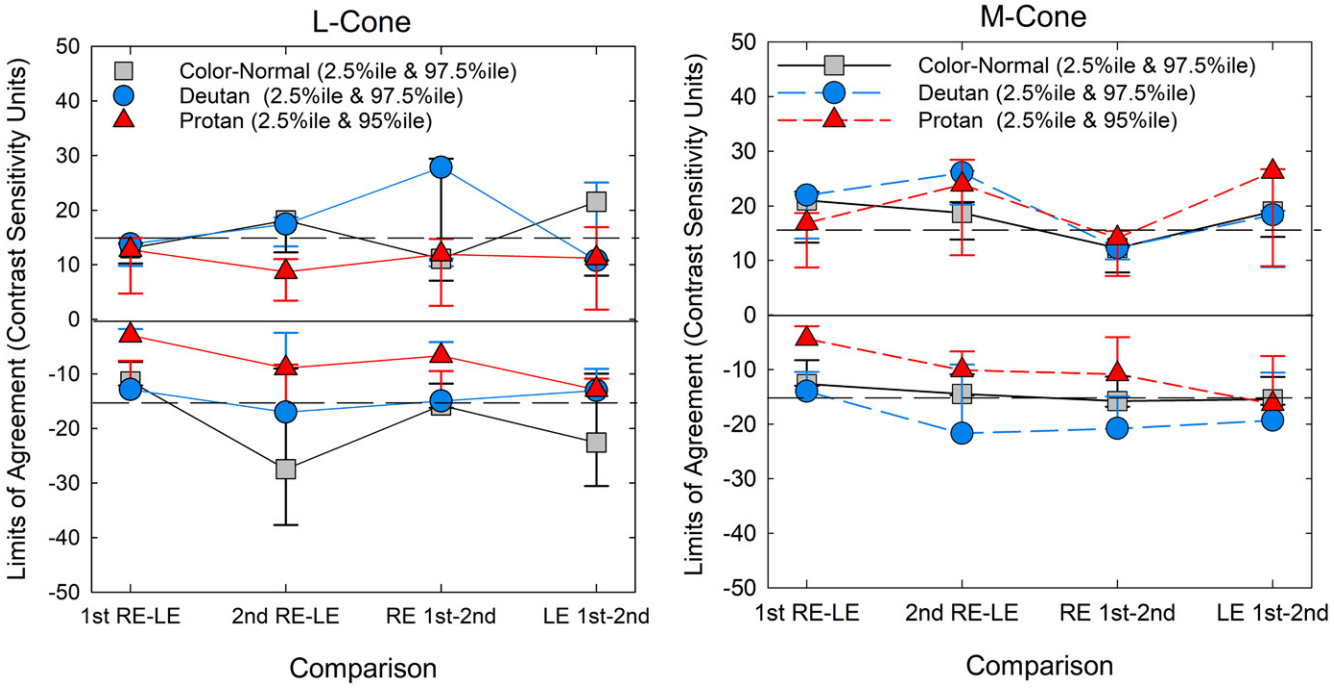

Fig. 6. OCCT limits of agreement for the various comparisons on the x-axis. 1st RE-LE is the between-eye comparison at the first session, 2nd RE-LE is the between-eye comparison at the second session, RE 1st–2nd is the comparison between first and second session for the right eye, and LE 1st–2nd is the comparison between the first and second session for the left eye. The dashed lines represent the COR of ±15 from previous studies.[4,8] Error bars represent the 95% confidence interval for the limits of agreement. Missing error bars indicate that the interval was zero. OCCT: Operational Based Visual Assessment Cone Contrast Test; COR: coefficient of repeatability.

the ±15 value. However, our LOA results from the M-cone sensitivity suggest that the between-eye and between-session values should be smaller at ±10. Whether the smaller value for the M-cone sensitivity LOA should be adopted may require more data to determine the tradeoff between the convenience of having the same LOA for the L- and M-cone vs. the ability to identify individuals with a monocular color vision loss. Our values for the NCV group are smaller than the ±30 value specified by Winterbottom et al., but their value is based on the combined results of NCV and DCV subjects.[9] Given that the LOAs for our results are greater for the DCV, it is not surprising that Winterbottom et al.'s values are larger when the DCV subjects are included with the NCV subjects.

In contrast to the RCCT, the OCCT LOAs for the NCV and deutans and the M-cone sensitivity for the protans were also reasonably close to ±15. On the absolute sensitivity scale, this corresponds to ±0.24 log units. A difference between the OCCT and RCCT LOAs is expected, particularly for the DCV groups. As Bailey et al.[15] pointed out, tests that use a coarser scale (e.g., RCCT) will have an artificially larger variation in their between-session results than those that use a finer scale (e.g., OCCT). This scaling effect is not seen with the NCV group because of the strong ceiling effect on the RCCT.

The asymmetry in the protan OCCT LOAs was due to the positive skewness of difference distributions. This result indicates that the larger between-eye and between-session differences are due to situations where the protan individual did worse on the second attempt. The reason for the smaller protan lower limit LOAs is not certain, but it could be related to the luminance contrast covarying with the contrast of the cone mechanisms. Although the stimuli are designed to isolate an individual cone mechanism, the stimuli do not have zero luminance contrast with the background, especially at the lower sensitivity values. Thus, the subject could potentially identify the gap based on chromatic contrast, luminance contrast, or both. We could not determine how the luminance contrast varied, but based on values reported for an early version of the RCCT, it is possible that the luminous contrast for the protan stimuli could have varied between 0.7% and 8%.[1] However, this range and the rate of change with the stimulus cone contrast would be different for the protan subjects since their luminous spectral sensitivity is reduced in the longer wavelengths and increased at medium wavelengths. How these differences interact with their threshold judgments is unclear, but it appears to limit any improvement in their performance on a second attempt, whether the second eye on the same day or a different session on another day.

Although the skewness in the protan L-cone differences might be related to their difference in spectral sensitivity and whether they can reliably interpret that information, Table I shows there were also L-cone differences for the NCV and deutans that were significantly different from a normal distribution, which suggests that there is something different with the OCCT L-cone measurement. These could be Type 1 errors, but the commonality across the three groups and the result that most of the M-cone differences did meet the normality assumption suggests otherwise. In all three cases where the NCV and deutan distributions did not meet the normality assumption, the distributions were leptokurtic; however, the deutans were positively skewed and the NCVs were negatively skewed. These results indicate that there were some individuals in both groups with large between-eye or between-session differences, but the deutan outliers were more likely to have lower sensitivity on the second session, and NCV outliers were more likely to have a higher sensitivity for the left eye, particularly at the second session. The reason for the nonnormal distributions in these two groups could also be related to interpreting luminance artifacts that may be present, or these results could be related to an order effect. The right eye was always measured first, but that would not easily explain the between-session differences, particularly the left eye between-session differences. The stimuli order was randomized within and between sessions, so one would expect that any within-session ordering effects would be minimized across subjects. If there is an order effect in the between-eye and between-session differences in the L-cone sensitivity, then it may become more apparent in the commercial version, which measures the L-cone sensitivity first, M-cone second, and S-cone third.

In addition to determining whether there has been a change in the person's color vision, the LOAs could be used to manage individuals who failed the test in one eye but passed in the other. In our study, none of the NCVs had between-eye pass/fail discrepancies at either failure criteria. However, 8% ($N = 5$) of DCV had these discrepancies on the RCCT at the first session using the ≥75 cutoff. In the second session, four out of five failed with both eyes or only with one eye. Of the 4 who failed at the second session, all 4 had between-eye differences at the first session that were greater than 15. The one individual who passed had a between-eye difference at the first session of 15. Thus, using the LOA of ±15 could help determine the color vision status of those who failed with one eye and passed with the other at the first session when using the RCCT. Whether this criterion would help resolve between-eye pass/fail discrepancies for the OCCT is uncertain because there were no between-eye discrepancies at the first session on the OCCT. The between-eye difference of the one individual who only passed in one eye at the second session was 5.2.

Using the ≥55 cutoff score for the RCCT, 14.3% ($N = 9$) had between-eye pass/fail discrepancies at the first session. The increase in discrepancies for the 55 score results from the large between-eye variability. Using the 15 value as the critical between-eye difference to resolve the between-eye discrepancy with a 55 pass/fail score on the RCCT is problematic because none of the NCV subjects had a score that low, and the LOA for the DCV group is considerably larger, particularly for the cone mechanism that is likely impaired in their visual system. If a difference in sensitivity of 15 is applied using the 55 cutoff score, then 4 individuals would pass on the first session, but only 1 would pass with each eye individually at the second session. The others would fail in either one or both eyes at the second session. Another option for the RCCT would be to use the

LOA for the cone mechanism with the greatest loss. That value would be approximately 40, averaged across the various comparisons. Although clinicians may be uncomfortable using 40 as an acceptable between-eye difference, it could serve as a helpful index. Two subjects had a between-eye pass/fail discrepancy and the difference between eyes was greater than 40 in the first session. Both failed the test in each eye in the second session. This result suggests that if there is a between-eye pass/fail discrepancy on the RCCT at the first session and the between-eye sensitivity differs by more than 40 units, then there is a reasonable probability that they will fail with each eye individually if tested on another day.

As with the RCCT, lowering the pass/fail cutoff to 55 increased the between-eye pass/fail discrepancies on the OCCT, but the total number was smaller. For the one deuteranomalous person with between-eye discrepancies at both sessions, the between-eye difference was approximately 21 at each session, which would be considered unacceptable based on the OCCT LOA for deutans. For the other 2 deuteranomalies (1 passed each eye at the first session and the other failed each eye at the first session), both also had between-eye differences at the second session of approximately 21. Because the number of discrepancies is small and none of the protan subjects had pass/fail discrepancies, it is difficult to determine how effective it is to use a between-eye difference of 15 to resolve between-eye discrepancies on the OCCT or whether a difference of 10 as found in this study should be applied to the protans.

Although the OCCT is a prototype and a newer version of this test and the RCCT are now available, the general findings of the RCCT ceiling effect and the differences in the RCCT LOA between NCV and DCV will likely remain. The larger LOA and lower repeatability of the RCCT relative to the OCCT could be due to other factors, such as differences in the staircase procedures, or if the RCCT letter legibility is not equivalent to the Landolt rings for chromatic stimuli. The major reason is likely the coarse contrast scale used by the RCCT vs. the finer scale used by the OCCT. Our results reiterate the advantages of using a finer scale to specify the stimulus magnitude in assessing the repeatability of clinical tests. Concerns have also been raised about the calibration of the RCCT and whether the contrast steps are evenly spaced.[9] While the advantage of the finer scale for determining whether the vision is normal vs. red-green defective may be only slight for test-retest repeatability, the finer scale does reduce the between-eye pass/fail discrepancies. The advantage of the finer scale becomes apparent when the pass/fail score is lowered and designed to allow some DCV to pass. In this case, the finer scale used in the OCCT resulted in a smaller between-eye and between-session variability and, thus, a higher test-retest repeatability.

ACKNOWLEDGMENT

Financial Disclosure Statement: This project was funded by Defence Research and Development Canada and the Canadian Institute for Military and Veteran Health Research (W7714-145967). Dr. Almustanyir was supported by a scholarship from the College of Applied Medical Science Research Center and the Deanship of Scientific Research at King Saud University. Dr. Hovis is a consultant for Alcon Laboratories. The authors have no competing interests to declare.

Authors and Affiliations: Jeffrey K. Hovis, OD, Ph.D., School of Optometry and Vision Science, University of Waterloo, Waterloo, Ontario, Canada; Ali Almustanyir, M.Sc., Ph.D., Optometry and Vision Sciences, Applied Medical Sciences College, King Saud University, Riyadh, Saudi Arabia; and Mackenzie G. Glaholt, MA, Ph.D., Human Effectiveness Section, Toronto Research Center, DGRDSE Defence Research and Development Canada, Toronto, Ontario, Canada.

REFERENCES

1. Rabin J. Cone-specific measures of human color vision. Invest Ophthalmol Vis Sci. 1996; 37(13):2771–2774.
2. Rodriguez-Carmona M, Harlow J, Walker G, Barbur J, editors. The variability of normal trichromatic vision and the establishment of the "normal" range. Proceedings of 10th Congress of the International Colour Association; May 8-13, 2005; Granada, Spain. Wien (Austria): Association Internationale de la Couleur; 2005.
3. Rodriguez-Carmona M, O'Neill-Biba M, Barbur JL. Assessing the severity of color vision loss with implications for aviation and other occupational environments. Aviat Space Environ Med. 2012; 83(1): 19–29.
4. Rabin J, Gooch J, Ivan D. Rapid quantification of color vision: the cone contrast test. Invest Ophthalmol Vis Sci. 2011; 52(2):816–820.
5. Walsh DV, Robinson J, Jurek GM, Capo-Aponte JE, Riggs DW, Temme LA. A performance comparison of color vision tests for military screening. Aerosp Med Hum Perform. 2016; 87(4):382–387.
6. Hovis JK, Almustanyir A. Assessment of the next generation of colour vision tests for pilots and aircrew. Toronto, Ontario (Canada): Defence Research and Development Canada; 2017. Report No.: DRDC-RDDC-2017-C19.
7. Gaska J, Winterbottom M, van Atta A. Operational based vision assessment cone contrast test: description and operation. Wright-Patterson AFB (OH): USAF School of Aerospace Medicine, Aeromedical Research Department; 2016. Report No.: FRL-SA-WP-SR-2016-0007.
8. Chay IW, Lim SWY, Tan BBC. Cone contrast test for color vision deficiency screening among a cohort of military aircrew applicants. Aerosp Med Hum Perform. 2019; 90(2):71–76.
9. Winterbottom M, Gaska J, Knapp J, Hadley S, O'Keefe E, et al. Comparison of cone contrast tests for color vision screening. Wright-Patterson AFB (OH): Air Force Research Laboratory, 711th Human Performance Wing, Airman Systems Directorate, Airman Biosciences Division, Performance Optimization Branch; 2019. Report No.: AFRL-RH-WP-TR-2019-0122.
10. Gao H, Kirkendall CD, Kinney MJ, Preston AM, Reddix MD. Color vision testing, standards, and visual performance of the U.S. military. Mil Med. 2023; 188(1–2):49–57.
11. Pokorny J, Smith VC, Verriest G, Pinckers AJLG. Congenital and acquired color vision defects. New York (NY): Grune & Stratton; 1979.
12. Kontsevich LL, Tyler CW. Bayesian adaptive estimation of psychometric slope and threshold. Vision Res. 1999; 39(16):2729–2737.
13. Lovell J, Rabin J. A comparison between three computer-based cone specific color vision tests. Aerosp Med Hum Perform. 2023; 94(2):54–58.
14. IBM SPSS Bootstrapping 28. Version 28.0.0. Armonk (NY): International Business Machines Corporation. [Accessed 2024 Nov. 13]. Available from https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Bootstrapping.pdf.
15. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. Invest Ophthalmol Vis Sci. 1991; 32(2):422–432.