AFRL-RH-WP-TR-2019-0122

COMPARISON OF CONE CONTRAST TESTS FOR COLOR VISION SCREENING

Marc Winterbottom, James Gaska, Jonelle Knapp, and Steve Hadley

711 HPW/RHBC OBVA Laboratory

Eleanor O'Keefe, Elizabeth Shoda, Eric Seemiller, and Alex Van Atta

KBR Inc.

July 2019
Final Report

Distribution Statement A. Approved for public release.

AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
AIRMAN SYSTEMS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2019-0122 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signature//
MARC WINTERBOTTOM, PhD
Senior Research Psychologist
Performance Optimization Branch
Airman Biosciences Division

//signature//
R. ANDY MCKINLEY, DR-III, PhD
Core Research Area Lead
Cognitive and Physical Performance
Performance Optimization Branch
Airman Biosciences Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
13-07-2019	FINAL	September 2018 – April 2019

4. TITLE AND SUBTITLE
Comparison of cone contrast tests for color vision screening

5a. CONTRACT NUMBER
FA8650-18-C-6932

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Marc Winterbottom, PhD, James Gaska, PhD, Eleanor O'Keefe, PhD, Elizabeth Shoda, M.S., Steve Hadley, MD, Eric Seemiller, PhD, Jonelle Knapp, OD, PhD, and Alex Van Atta

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
Legacy RHM

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
OBVA Laboratory
Airman Biosciences Division
Airman Systems Directorate
711th Human Performance Wing
WPAFB, OH 45433

KBR
Space and Mission Solutions
2601 Mission Point Blvd., Suite 100
Beavercreek, OH 45431

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
711th Human Performance Wing
Airman Systems Directorate
Airman Biosciences Division
Performance Optimization Branch
Wright-Patterson AFB, OH 45433

10. SPONSORING/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RH-WP-TR-2019-0122

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution Statement A. Approved for public release.

13. SUPPLEMENTARY NOTES
Report contains color. AFRL-2021-3506, cleared 26 October 2021

14. ABSTRACT
Good test-retest reliability is essential to predict relevant performance (e.g. flying performance) and maintain sensitivity/specificity. The cone contrast test (CCT) has been used since 2011 for aircrew color vision screening. However, there is very little test-retest reliability data. This research compares the color calibration and test-retest reliability of four versions of the CCT: 1) Innova Systems, v18 (New Innova), 2) Innova Systems, v13.11 (Old Innova), 3) NCI Vision Systems, v14 (NCI), and 4) Konan Medical CCT-HD, v1.0.0.1 (Konan). Monocular CCT scores were measured for L-, M-, and S-cones for 67 subjects. Each subject repeated each test 4 times: (1) right eye (OD) day 1; (2) left eye (OS) day 1; (3) OD day 2; and (4) OS day 2. We used a Bland-Altman analysis to compare the 95% limits of agreement (LOA) for Test 1 and Test 2 for each of the 4 tests and for each eye. Based on the cone contrast measurement results, the Konan Medical CCT HD clearly has the most precise calibration. In contrast, the difference between measured and desired contrast for the Innova and NCI CCTs is much larger on average. Results also revealed that LOAs were lowest (i.e. best test-retest reliability) for the Konan L- and M-cone tests. For the S-cone, the Konan had the largest LOAs. However, there was a large ceiling effect for the S-cone test for both the Innova and NCI tests. Although all of the CCTs have been shown to reliably screen for color deficiency (for pass/fail criterion of 75), only the Konan CCT-HD provides precision color calibration, high test-retest reliability for L- and M-cones, and can fully characterize both color deficient and color normal individuals to support Air Force Medical Service objectives to develop better, more diagnostic, threshold level aeromedical tests.

15. SUBJECT TERMS
Human vision system, operational performance, contrast sensitivity, visual discrimination

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	SAR	43	Marc Winterbottom
					19b. TELEPHONE NUMBER *(include area code)* n/a

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF EQUATIONS

1.0 SUMMARY

Good test-retest reliability is essential to predict relevant performance (e.g. flying performance) and maintain sensitivity/specificity. The cone contrast test (CCT) has been used since 2011 for aircrew color vision screening. However, there is very little test-retest reliability data. This research compares the color calibration and test-retest reliability of four versions of the CCT: 1) Innova Systems, v18 (New Innova), 2) Innova Systems, v13.11 (Old Innova), 3) NCI Vision Systems, v14 (NCI), and 4) Konan Medical CCT-HD, v1.0.0.1 (Konan). Monocular CCT scores were measured for L-, M-, and S-cones for 67 subjects. Each subject repeated each test 4 times: (1) right eye (OD) day 1; (2) left eye (OS) day 1; (3) OD day 2; and (4) OS day 2. We used a Bland-Altman analysis to compare the 95 percent (%) limits of agreement (LOA) for Test 1 and Test 2 for each of the 4 tests and for each eye. A reliable test will result in a narrower LOA. Based on the cone contrast measurement results, the Konan Medical CCT HD clearly has the most precise calibration, with an average percent error less than 10% for each cone. In contrast, the difference between measured and desired contrast for the Innova and NCI CCTs is much larger on average, in some cases with errors exceeding 50%, particularly for low contrast stimuli. Averaged across OD and OS, the LOAs for the Konan, New Innova, Old Innova, and NCI were, respectively, 19.1, 21.1, 29.3, and 30.3 for the L-cone; 21.4, 23.6, 28.9, and 30.8 for the M-cone, and 35.6, 20.1, 21.2, and 13.4 for the S-cone. Thus, LOAs were lowest (i.e. best test-retest reliability) for the Konan L- and M-cone tests. For the S-cone, the Konan had the largest LOAs. However, there was a large ceiling effect for the S-cone test for both the Innova and NCI tests, with approximately 50% of subjects scoring at the maximum value of 100 for both Innova CCTs, and 84% of subjects scoring at the maximum value of 100 on the NCI CCT. Although all of the CCTs have been shown to reliably screen for color deficiency (for pass/fail criterion of 75), only the Konan CCT-HD provides precision color calibration, high test-retest reliability for L- and M-cones, and can fully characterize both color deficient and color normal individuals to support Air Force Medical Service objectives to develop better, more diagnostic, threshold level aeromedical tests. The results of this research suggest that additional research is needed to examine the utility of S-cone testing.

2.0 INTRODUCTION

Color vision capabilities, like other visual capabilities, are characterized by individual differences even within the normal population. However, because color vision deficiency (CVD) can profoundly decrease color discrimination in affected individuals (Cole, 2004) there has been a long history of identifying CVD individuals and excluding them from occupations that require critical decisions that may depend on color vision. The importance of occupational color vision screening has been recognized since at least 1875, following a serious train crash which was attributed to a failure to recognize signal lights (Fee, 1879; Verriest, 1979). Given considerable evidence that color vision deficient observers perform more poorly than color vision normal (CVN) observers in color discrimination tasks (Cole, 1993, Cole, 2004, Steward & Cole, 1989) it is apparent that selection and retention standards related to color vision are necessary for aircrew. However, they should be evidence-based standards supported by data showing the relationship between color vision and relevant operational task performance.

Normal color vision is supported by three retinal cone types, with peak spectral sensitivities in the long (L), middle (M) and short (S) wavelength sections of the visible spectrum. In addition to the L, M and S nomenclature, these three cone types are often referred to as the red, green and blue cones, respectively. Because there are three cone types, normal color vision is labeled trichromatic. Color deficiency is first identified by the type of cone affected: protan (L, red), deutan (M, green) and tritan (S, blue) and then by the severity of the deficiency. The most severe form of color deficiency is caused by the absence of one cone type (leaving two) and is called dichromacy. There are three types of dichromacy (identified by cone type): protanope, deuteranope and tritanope. Protanopia and deuteranopia are sex-linked and can affect as many as 5-8 percent (%) of European descendent males and < 0.5% of European descendent females, with much lower incidence in Asian, African and Native American populations (Delpero, O'Neill, Casson & Hovis, 2005). Tritanopia is not sex-linked and considerably rarer, affecting less than 0.01% of the population (Wright, 1952). Individuals with a less severe and more common form of CVD (anomalous trichromacy) retain the full complement of three cones; however, the spectral sensitivity of one of the cones differs from CVN individuals. Again, there are three forms of anomalous trichromacy: protanomalous, deuteranomalous and tritanomalous.

Although there are genetic tests that can identify if an individual has missing (dichromats) or abnormal (anomalous trichromats) cone photopigments, these techniques are not capable of precisely measuring color vision capabilities, which also depend on optical (pre-retinal and macular absorption) and neural factors. For many years, color vision screening was carried out primarily using pseudoisochromatic plates (PIP) tests, such as the Ishihara test (Ishihara, 1918). However, motivated individuals can prepare for and pass the PIP tests despite having color vision deficiencies. Additionally, the PIP tests cannot reliably indicate the severity of color deficiency (Barbur & Rodriguez-Carmona, 2017). Fortunately, several computer-based, automated tests have recently been developed that reduce/eliminate many of the disadvantages of the PIP tests.

The Rabin Cone Contrast Test (RCCT) was developed by the United States Air Force (USAF) to replace PIP tests, which had been in use for many years to screen for CVD (Rabin, 1996; Rabin, 2004; Rabin, Gooch, & Ivan, 2011). The USAF adopted the RCCT in 2011 for aircrew vision screening and it has been shown to reliably screen for CVD (Rabin et al, 2011; Hovis & Almustanyir, 2017). There are now two versions of the RCCT available on the commercial market – the Innova Systems RCCT (www.innnovasystemsusa.com) and the NCI Vision Systems OcuTest Extended with the RCCT (www.ncivision.com/ocutest-extended-1).

The USAF Operational Based Visual Assessment (OBVA) Laboratory developed the next generation cone contrast test, the OBVA CCT, or OCCT, to overcome some limitations of the earlier generation of CCTs and to support (1) research that quantifies the relationship between color vision and operational performance and (2) evidence-based color vision standards. The OCCT was developed with improved precise color calibration; adoption of a validated adaptive threshold estimation procedure (Psi method, Prins, 2018); improved test-retest reliability (specificity); and a simplified four-alternative forced choice response using a Landolt C optotype. The OCCT, with increased precision in color calibration, can achieve the very low cone contrast values needed to accurately quantify cone contrast thresholds even for CVN individuals. The capability to test CVN individuals is essential in order to research the relationship between color vision and operationally-relevant performance. It also enables early detection of disease, improved ability to quantify the effect of hypoxia and other environmental stressors on vision, and the investigation of the potential effects of chemicals/pharmaceuticals.

The development of the OCCT has been described in more detail in a previous report (Gaska, Winterbottom, & Van Atta, 2016). In 2016, the OBVA Laboratory partnered with Konan Medical through a cooperative research and development agreement (CRADA) to develop a commercial version of the OCCT – the Konan Medical ColorDx CCT HD (https://www.konanmedical.com/colordx). Research conducted to validate the CCT HD is described in more detail in a previous report (Gaska, Winterbottom, Hadley, O'Keefe, Shoda, & Van Atta, 2021).

The purpose of the research summarized in this paper was to evaluate and compare several different versions of the CCT. Each of these tests are variants of the CCT originally developed for the U.S. Air Force (USAF) by Jeff Rabin (Rabin, 1996; Rabin, Gooch, & Ivan, 2011). The tests evaluated included:

1. Innova Systems RCCT, version 18 (referred to as New Innnova)
2. Innova Systems ProVideo RCCT, version 13.11 (referred to as Old Innova)
3. NCI Vision Systems OcuTest Extended RCCT, version 14 (referred to as the NCI)
4. Konan Medical ColorDx CCT HD, version 1.0.0.1 (referred to as the Konan)

3.0 METHODS

3.1 CCT Calibration

All color tests were calibrated as per manufacturer guidelines using the built-in calibration software. The NCI and New Innova systems were calibrated with a Datacolor Spyder 5 colorimeter (Datacolor, Lucerne, Switzerland), the Old Innova system with a Datacolor Spyder 4 colorimeter, and the Konan with an X-rite i1 Display colorimeter (X-rite, Grand Rapids, MI) as shown in Figure 1 (left).

3.2 Cone Contrast Measurement

For the Konan evaluation, the desired contrast condition was varied over the range of 0.001 to 1.0 in 0.1 log unit steps in low to high order and 3 cone isolations in order (Long, Medium, and Short wavelength, or L, M and S). Each cone isolation was measured at each contrast before proceeding to the next higher contrast. This block of measurements was repeated 21 times and averaged over time. The total measurement time was approximately 8 hours.

We could not directly control the contrast of the Innova and NCI CCT devices. Instead, we ran the test program and used a screen capture to record the r, g and b values for the five contrast levels used in the tests. We then used the red, green, blue (RGB) values of the stimulus and background to create test patches to measure the L, M and S contrast values (described below). Two measurements were made for every test image and the results were averaged. The nominal or desired contrast of these test stimuli are provided in Table 1. Note that the L and M values are identical, and the S values are approximately a factor of ten higher than the L and M values.

Table 1. Desired contrast values for Rabin CCT

Desired Contrast Values for Rabin CCT					
Contrast			Log10(Contrast)		
L	M	S	Log10(L)	Log10(M)	Log10(S)
0.200	0.200	1.585	-0.7	-0.7	0.2
0.112	0.112	0.891	-0.95	-0.95	-0.05
0.063	0.063	0.501	-1.2	-1.2	-0.3
0.022	0.022	0.282	-1.65	-1.65	-0.55
0.013	0.013	0.158	-1.9	-1.9	-0.8

3.3 Measuring LMS Contrast

An Ocean Optics Maya 2000 Pro spectrophotometer was used to measure the spectral irradiance of light generated by the stimulus and background for each contrast level (see Figure 1, right). The dot product between the spectral irradiance and the three (L, M and S) Stockman and Sharpe (2000) 2° cone fundamentals was used to compute L, M and S cone excitation levels for both background and stimulus. Cone contrast values were computed as the difference between the cone excitation of the stimulus and background divided by the background excitation ($\Delta L/L$, $\Delta M/M$ and $\Delta S/S$), where delta (Δ) is the difference between the stimulus and the background for each cone as shown in Equation 1 below.

$$contrast = \frac{(stimulus\ excitation - background\ excitation)}{background\ excitation}$$

Equation 1. Cone contrast equation.

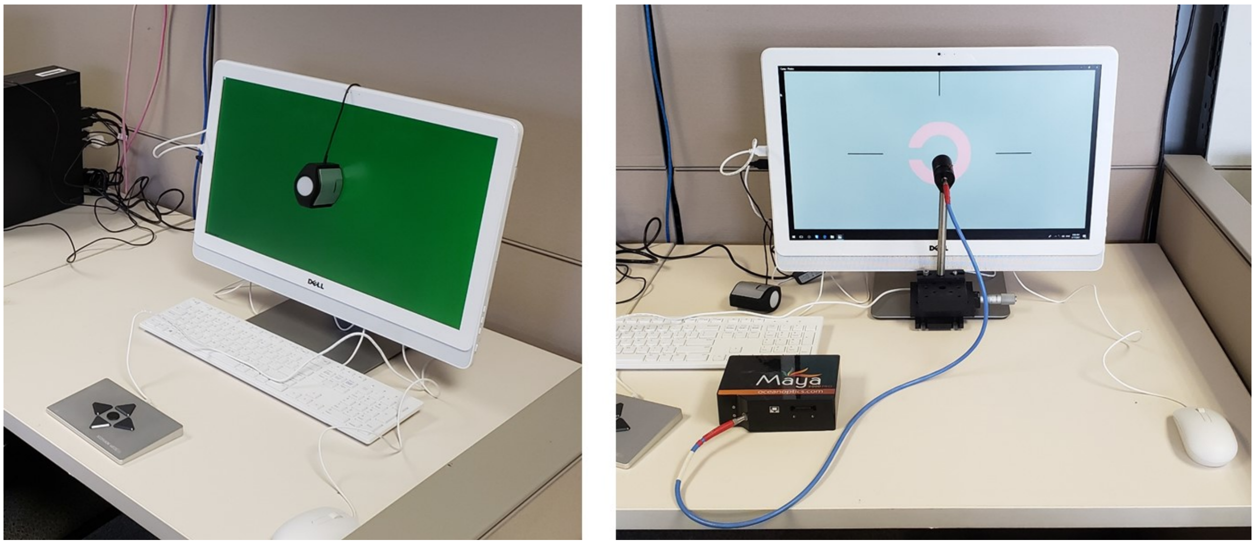

Figure 1. Konan CCT HD with X-rite i1 Display colorimeter (left) and with Ocean Optics Maya 2000 Pro spectrophotometer (right)

3.4 Cone Contrast Measurement Results

Figures 2–5 show the calibration test results for all devices. The left column of the figures plots the desired cone contrast of the isolated cone on the horizontal axis and the measured cone contrast of the isolated cone on the vertical axis. L, M and S results are plotted using red, green and blue symbols, respectively. The right column of the figure plots a vector error metric as a function of desired cone contrast. The error metric, E, is defined in Equation 2 below.

$$E = 100 \cdot D/C_{iso}$$

Equation 2. Cone contrast error metric equation.

Where the Euclidian distance, D, between the desired cone contrast and the measured cone contrast in 3-dimensional cone contrast space and normalizing the distance by the desired cone contrast of the isolated cone type, C_{iso}, as defined in Equation 3 below. Multiplying by 100 converts the proportional error to percent error.

$$D = \sqrt{(L_{measured} - L_{desired})^2 + (M_{measured} - M_{desired})^2 + (S_{measured} - S_{desired})^2}$$

Equation 3. Vector distance equation between the measured and desired cone contrast.

Figure 6 shows the desired cone contrast and vector error metric for each test for the same cone on the same plot for more direct comparison of the precision of the calibration for each test. This figure shows more clearly that the Konan CCT HD is capable of displaying the very low contrasts needed to estimate threshold level contrast for CVN individuals – more than a log unit lower contrast than either the Innova or NCI CCTs. Figure 6 also clearly shows the precise calibration of the Konan CCT HD even at very low levels of contrast.

Figure 7 summarizes the mean calibration error for each CCT. As shown, the Konan Medical CCT HD clearly has the most precise calibration, with an average percent error less than 10% for each cone. In contrast, the difference between measured and desired contrast for the Innova and NCI CCTs is much larger on average, in some cases with errors exceeding 50%, particularly for low contrast stimuli as shown in Figures 3 - 5.

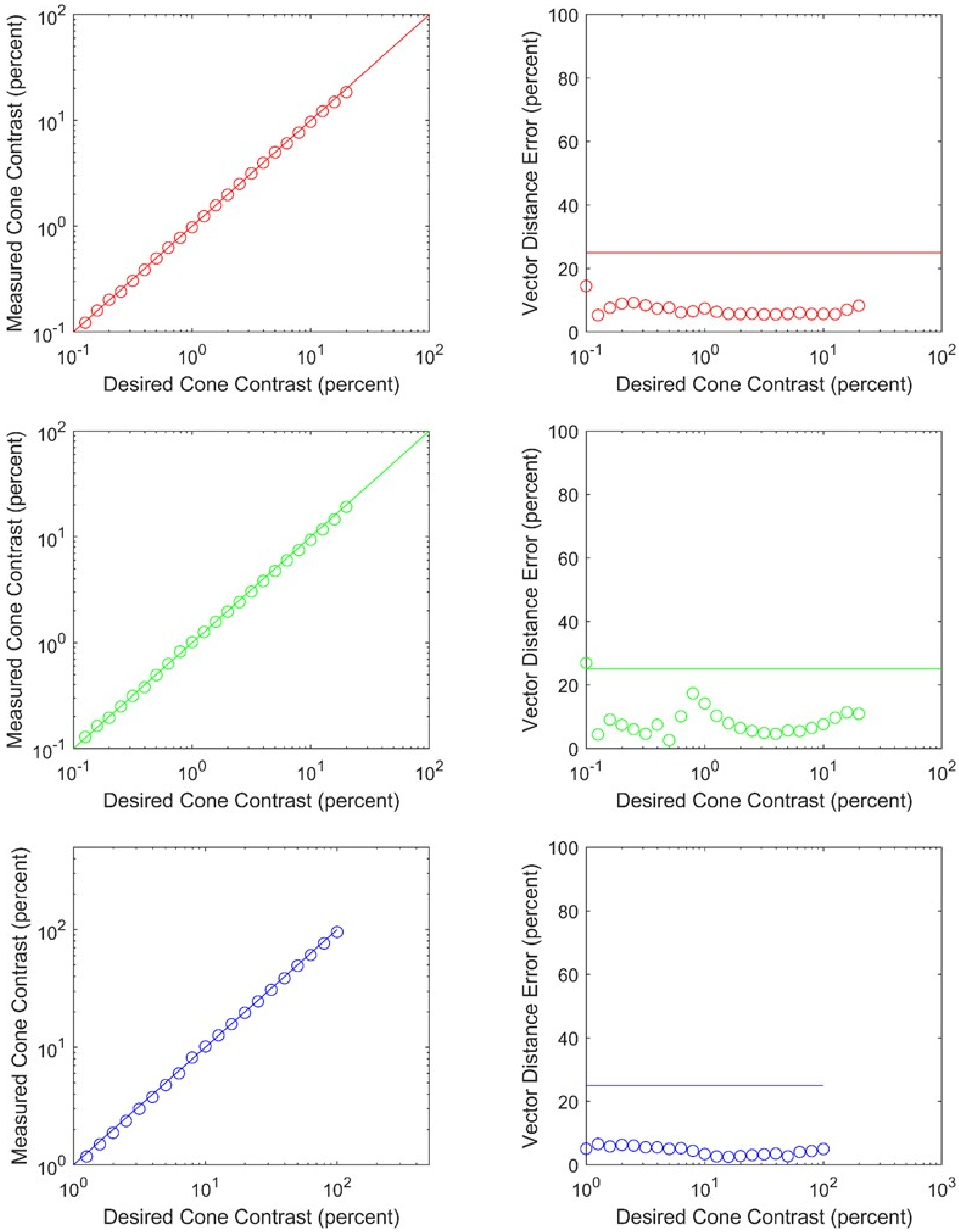

Figure 2. Cone contrast measurement results for the Konan

Left column: desired vs. measured cone contrast. Right column: Vector error metric as a function of desired cone contrast.

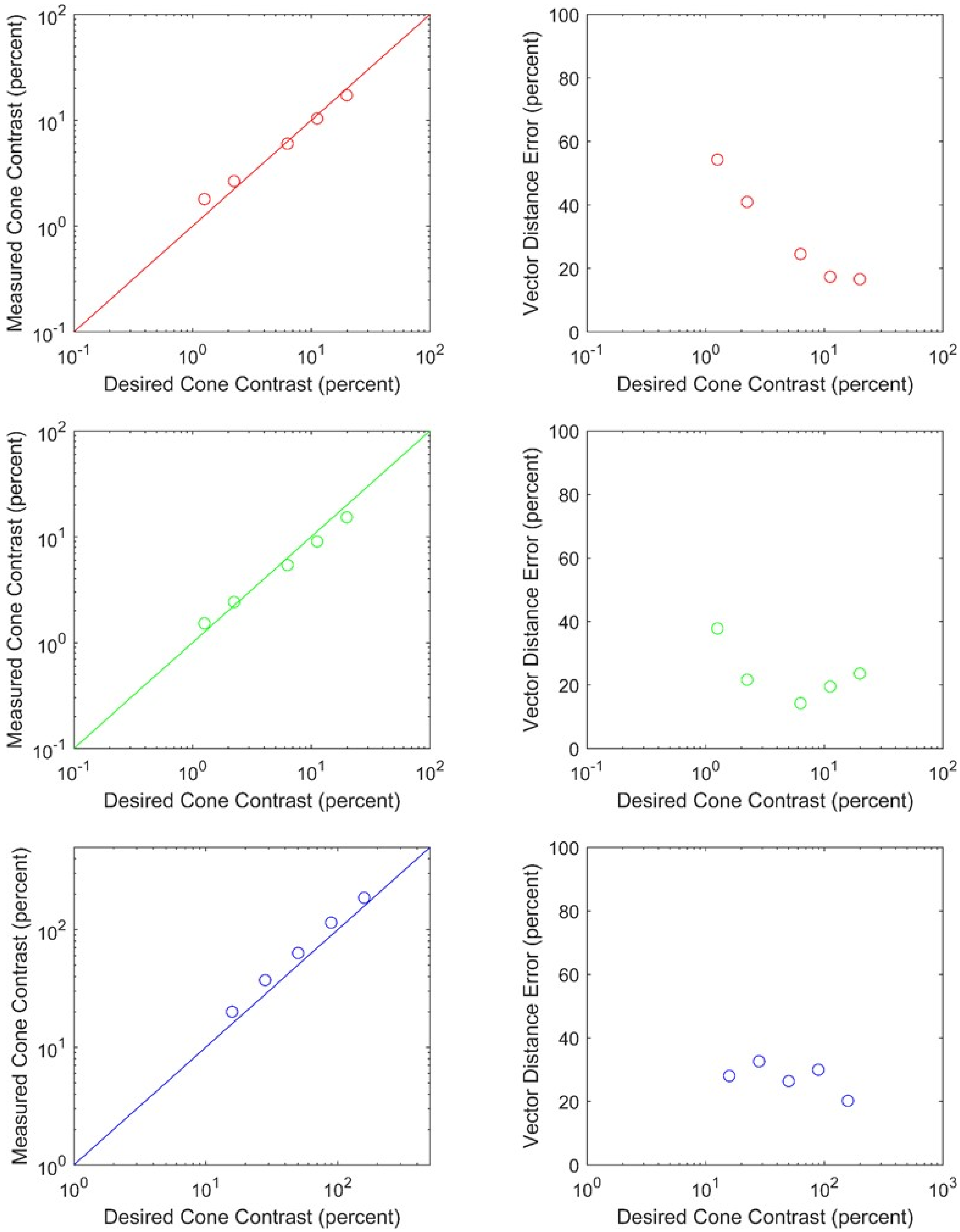

Figure 3. Cone contrast measurement results for the Old Innova

Left column: desired vs. measured cone contrast. Right column: Vector error metric as a function of desired cone contrast.

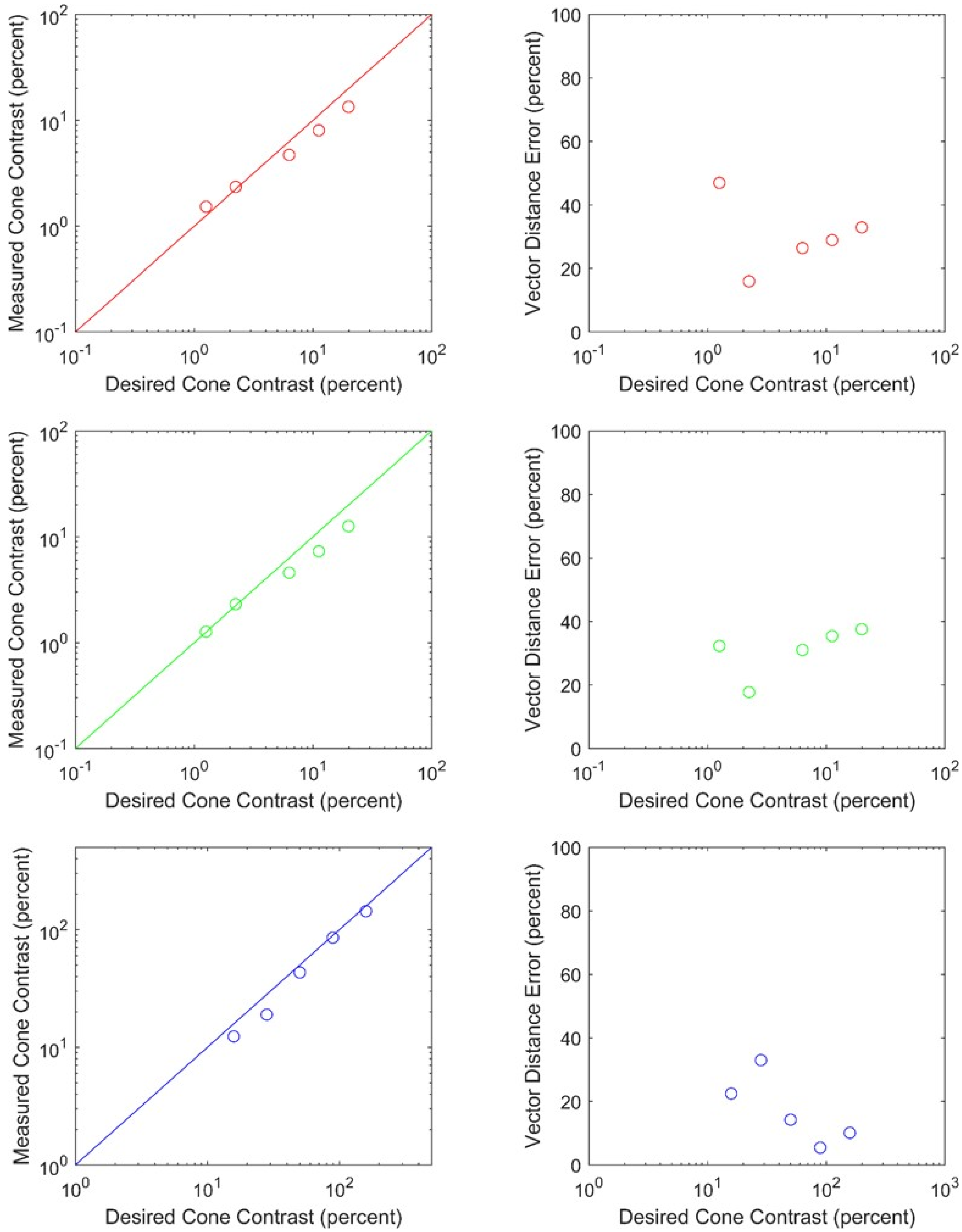

Figure 4. Cone contrast measurement results for the New Innova

Left column: desired vs. measured cone contrast. Right column: Vector error metric as a function of desired cone contrast.

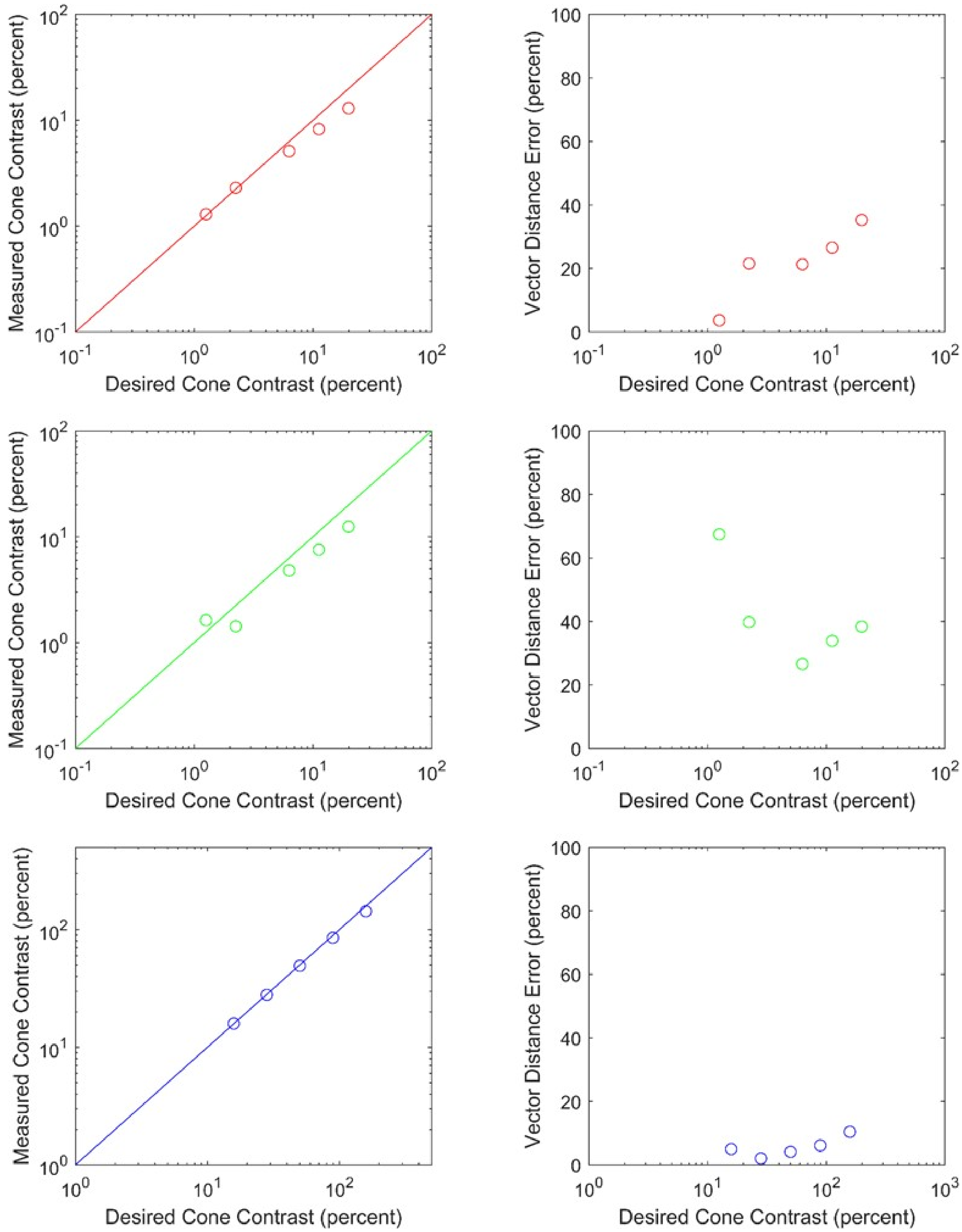

Figure 5. Cone contrast measurement results for the NCI

Left column: desired vs. measured cone contrast. Right column: Vector error metric as a function of desired cone contrast.

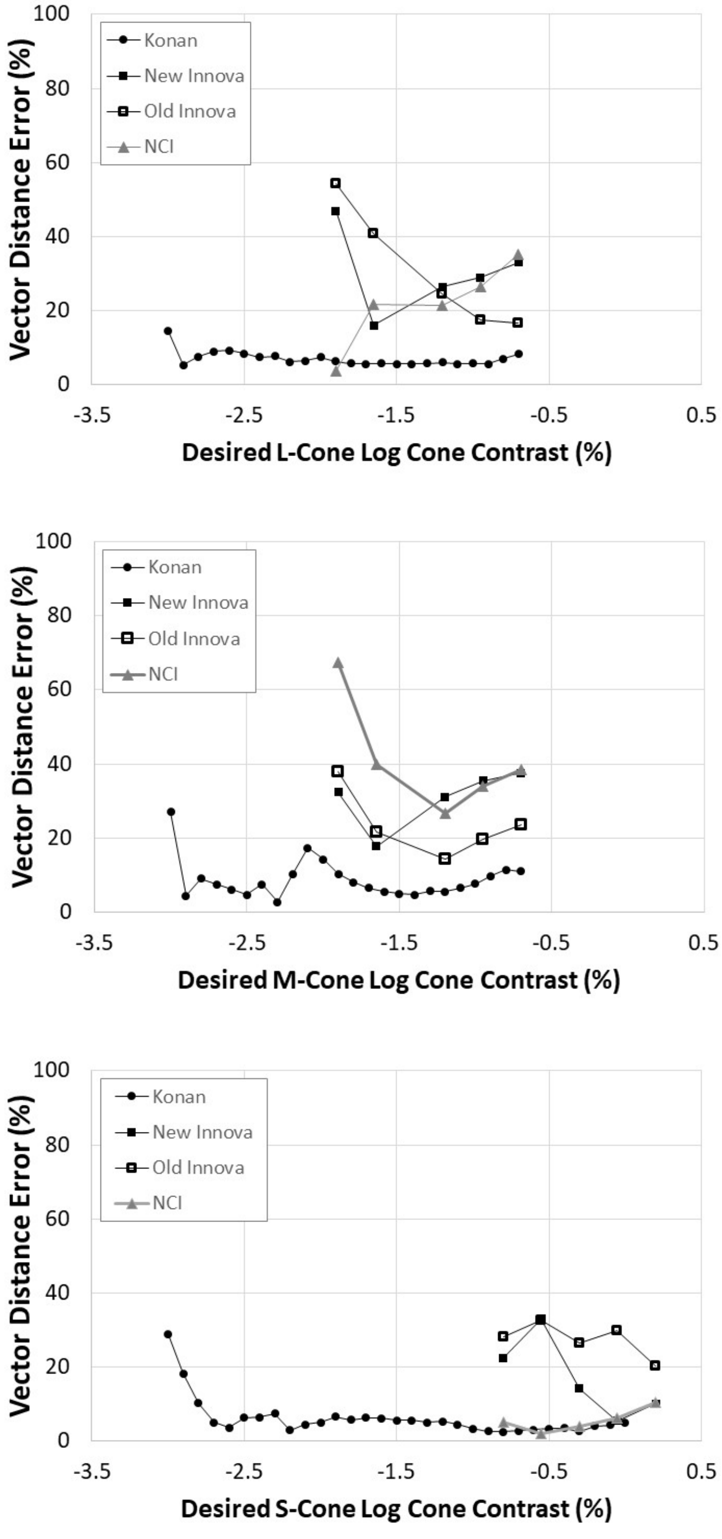

Figure 6. Desired contrast and vector error metric for L-cone (top), M-cone (center) and S-cone (bottom) for each test

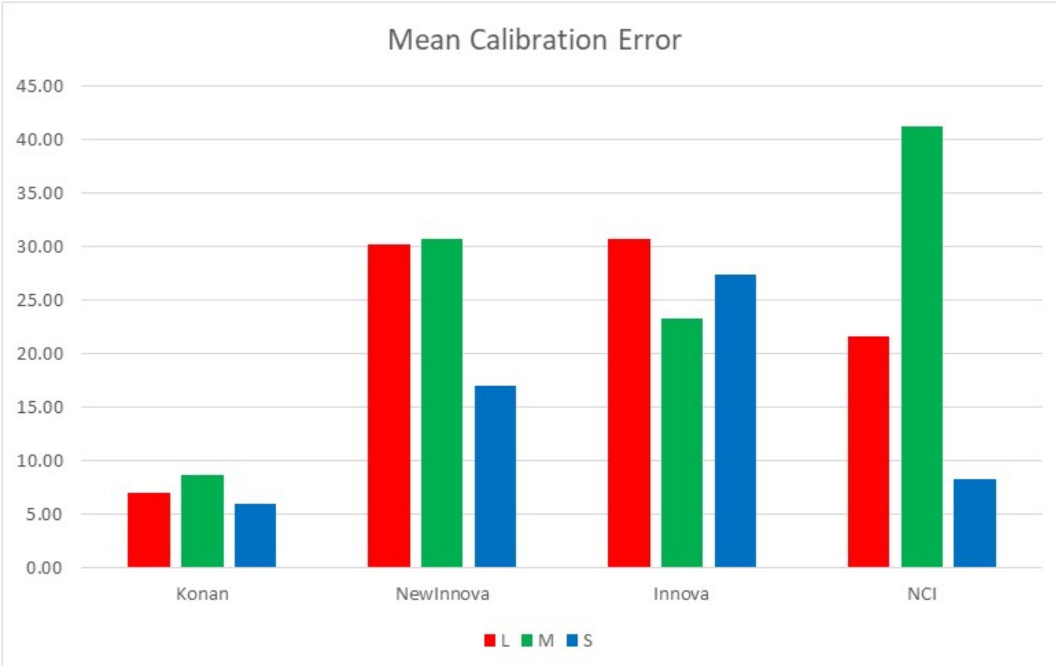

Figure 7. Summary of average calibration error for each CCT

4.0 TEST-RETEST RELIABILITY

For a valid, accurate and repeatable test, an individual tested on one day should produce approximately the same test score when tested on a different day. Tests with low between-day score differences have good test-retest reliability. We tested 67 observers on two different days to quantify test-retest reliability of each of the CCTs. Observers were tested monocularly using an eye patch. Thus, each observer repeated each test 4 times: (1) Oculus Dexter (right eye) (OD) day 1; (2) Oculus Sinister (left eye) (OS) day 1; (3) OD day 2; and (4) OS day 2. Observers were screened for color deficiency prior to participating in the study in order to recruit protanomalous and deuteranomalous observers. The study was conducted in accordance with Air Force Research Laboratory Institutional Review Board protocol number FWR20170095H v2.01.

4.1 Methods

There were several variable settings for each test including observer distance, optotype size, stimulus presentation time and the option to give additional response time. All settings were followed according to manufacturer guidelines except for the additional response time in the New Innova and the NCI. These were set in order to be comparable to the Konan as well as to provide adequate response time. However, the Old Innova did not provide the option to adjust response time. Table 2 summarizes the configuration of each of the 4 CCTs used to examine test-retest reliability.

Table 2. Configuration of each CCT

	Observer distance (m)	Optotype size (height)				Stimulus presentation time (s)	Additional Response Time (s)
		L and M Cone Stimulus	S Cone Stimulus	Bank of Letters	Mouse		
Old Innova	0.91	20/377	20/453	20/151	20/76	4	0
New Innova	1	20/344	20/412	20/138	20/69	6	14
NCI CCT	1	20/344	20/440	N/A	N/A	6	14
Konan CCT	1	20/344	20/344	N/A	N/A	8	unlimited

4.2 Old Innova

The Innova Systems' ProVideo Rabin CCT, version 13.11, referred to as "Old Innova" in this study, used an Acer Aspire V5-131 Series laptop (Acer, Taipei, Taiwan) with participant responses entered using a mouse. The Old Innova RCCT uses Bailey-Lovie letters (H, N, V, R, U, E, D, F, P, Z). Participants were asked to identify the letter displayed using the mouse and a response grid in the bottom right corner of the screen (see Figure 8). One of the response options was "Pass", but participants were instructed not to select this option and instead provide their best guess if they were uncertain as to which letter was displayed. The first letter presented for each color was the highest cone contrast and a staircase procedure using five contrast levels determined the participant's sensitivity (not the full letter chart scoring procedure described by Rabin, Gooch, and Ivan, 2011). Sensitivity scores ranged from 0 to 100. The staircase procedure is a customized procedure described in the product manual. Figure 9 (left) shows a participant completing the Old Innova test.

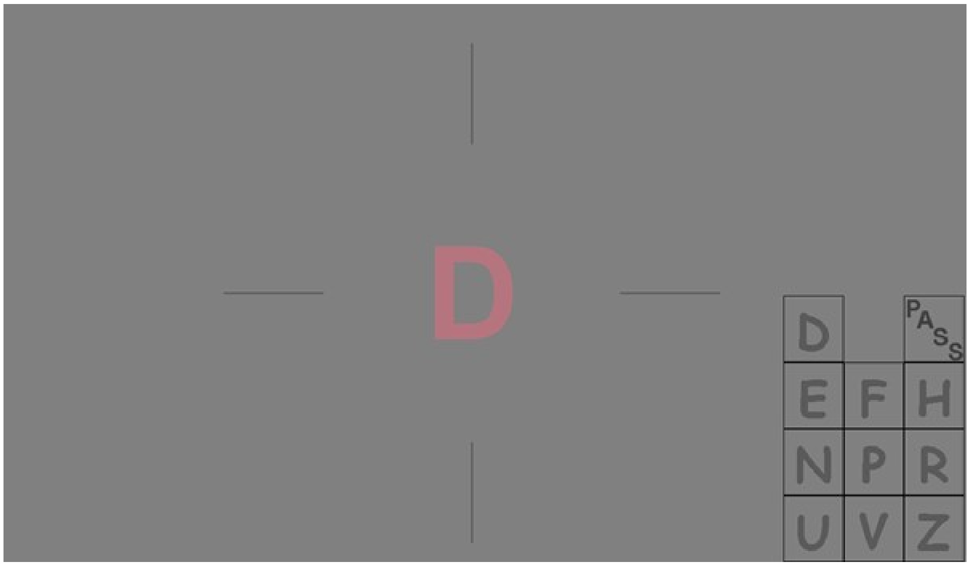

Figure 8. Old Innova optotype and response selection

4.3 New Innova

The Innova Systems' Rabin Cone System, version 18, referred to as the "New Innova" in this study, used a Microsoft Surface Pro tablet (Microsoft, Redmond, Washington). The New Innova uses a higher quality display, but otherwise is similar to the Old Innova. It uses the same Bailey-Lovie letters (H, N, V, R, U, E, D, F, P, Z) and participants were asked to identify the letter displayed using the mouse and a response grid in the bottom right corner of the screen (see Figure 8). As with the Old Innova, one of the response options was "Pass", but participants were instructed not to select this option and instead provide their best guess if they were uncertain as to which letter was displayed. The first letter presented for each color was the highest cone contrast and a staircase procedure using five contrast levels determined the participant's sensitivity. Sensitivity scores ranged from 0 to 100. Figure 9 (right) shows a participant completing the New Innova test while wearing an eye patch.

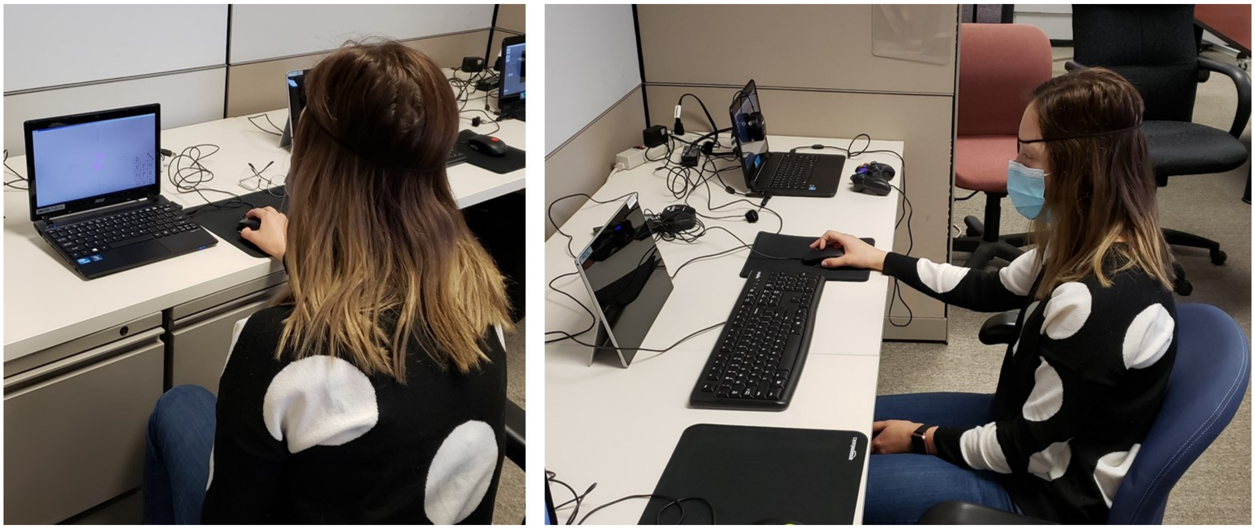

Figure 9. Participant completing the Old Innova test (left) and the New Innova test (right)

4.4 NCI

The NCI Vision Systems OcuTest Extended RCCT, version 14, referred to as the NCI in this study, used a Lenovo N23 laptop (Lenovo, Beijing, China) with participant responses entered using an EasySMX wired controller (EasySMX, Shenzhen, China). The NCI also employs the Rabin CCT Stair Test and uses the same five contrast levels used by the Innova CCTs. However, instead of optotype selection using a mouse, the NCI uses a four-alternative forced-choice design with a Landolt C stimulus. In each trial, the Landolt C appeared in the center of the screen with the gap in the C facing one of four possible orientations: up, down, right, or left (Figure 10). Participants were instructed to use the x, y, a, and b buttons on the controller to select the direction of the gap in the C. As with the Innova version of the Rabin CCT, sensitivity scores ranged from 0 to 100. Figure 10 shows a participant using the game controller to complete the NCI test.

Figure 10. Participant using the game controller to complete the NCI test

4.5 Konan

The Konan Medical ColorDx CCT HD, version 1.0.0.1, referred to as the “Konan CCT” in this study, used a Dell Inspiron 22 all-in-one computer (Dell, Round Rock, TX) with participant responses entered using a proprietary 4-direction response pad (see Figure 12, left). The Landolt C and four-alternative forced-choice procedure was based on the OCCT (see Figure 11) as described in a previous technical report (Gaska et al, 2016). Participants used the response pad to indicate the direction of the gap in the C. Unlike the other CCTs, the Konan uses the Psi algorithm (Prins, 2018) to calculate the participant’s threshold in 32 trials.

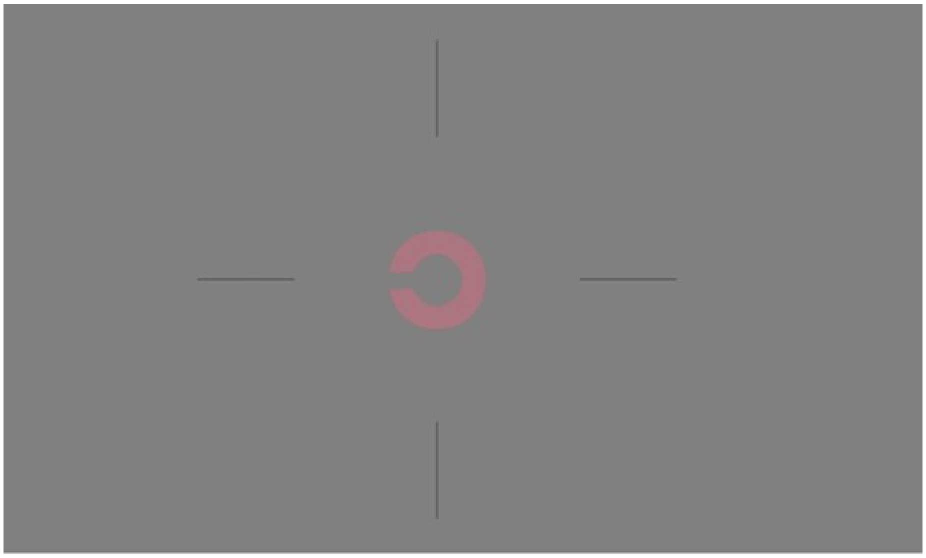

Figure 11. Konan Landolt C optotype (L-cone)

Auditory feedback was also provided, with a high tone indicating a correct response, and a low tone indicating an incorrect response. Cone contrast threshold scores were given in log cone contrast. Unlike the other CCTs, which display only 5 contrast levels, the Konan is capable of displaying a much wider range of cone contrast levels as shown in Figures 2 and 6. Figure 12 shows a participant seated at the Konan CCT HD.

Figure 12. Konan CCT HD response pad (left) and research participant completing the Konan CCT HD test (right)

To facilitate comparison between the tests, the Konan log cone contrast threshold was converted to a score using the equation below. Additionally, because of the precise calibration for the Konan, a wider range of scores is achievable, ranging from 0 to 140, rather than 0 to 100 as with the other CCTs.

$$Score = (KonanLogCS * 100) - 90$$

Equation 4. Conversion from Konan log cone contrast to a score.

Although it is not relevant to this study focusing on test-retest reliability, and scores are not compared directly between tests, it should be noted that the Konan scores derived from Equation 4 are not directly comparable to the other CCTs. This is because the relationship between cone contrast and CCT score is non-linear for the Innova and NCI CCTs as shown in Figure 13.

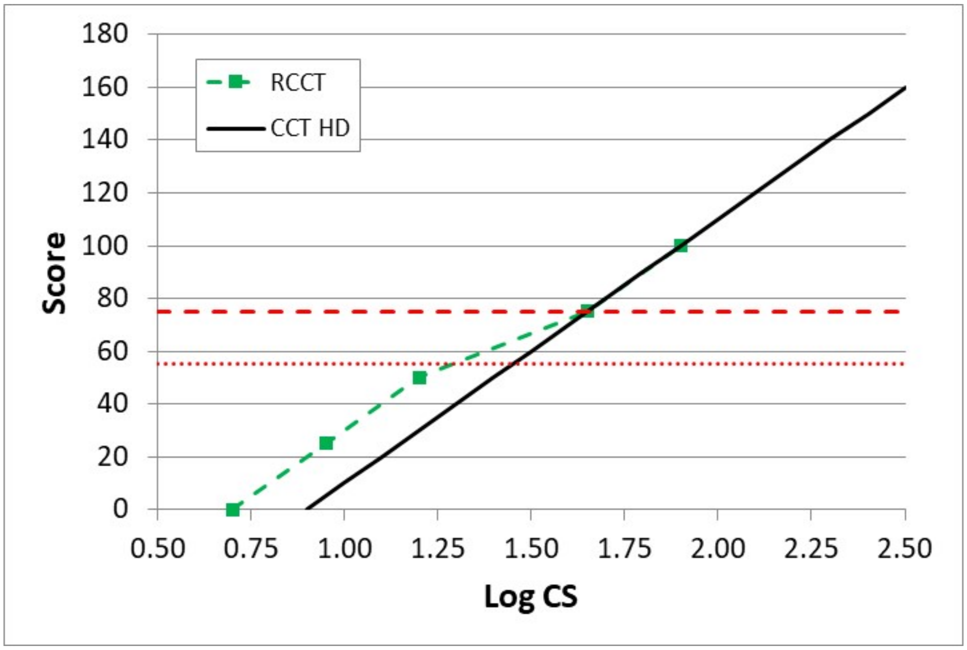

Figure 13. Relationship between log cone contrast and CCT score for the RCCT (NCI and Innova, green symbols) and the Konan CCT HD (black line).

The green symbols show the 5 contrast levels used by the NCI and Innova. The red dashed line shows the USAF pass/fail criterion of 75 used until 2018 and the red dotted line shows the current USAF and USN pass/fail criterion of 55.

4.6 Analysis

Because the highest achievable score for the Innova and NCI RCCTs is 100, many CVN individuals will simply score the maximum value of 100 on both days. Because of this ceiling effect, these scores would generate artificially low between-day score differences (0). To eliminate the effect of the CCT score ceiling effect, any test that resulted in a score of 100 was excluded from the test-retest analysis as well as its inter-day pair member.

We used a Bland-Altman analysis (Giavarina, 2015) to examine the differences between test 1 (T1) and test 2 (T2). They show the mean score [(T1+T2)/2] on the horizontal axis and the difference score (T1-T2) on the vertical axis. The plots also show the upper and lower 95% Limits of Agreement (LOA), which is the mean difference plus or minus 1.96 times the standard deviation of the differences. A highly reliable test will result in a mean difference between T1 and T2 near zero, and narrower LOA.

4.7 Results

The percent of inter-day pairs that resulted in a ceiling effect is shown in Table 3. The highest percentages are for the S cone, particularly for the NCI device.

Table 3. Percent of inter-day test-retest pairs that resulted in a ceiling effect

	L	M	S
Konan	0%	0%	0%
New Innova	24%	6%	47%
Old Innova	4%	47%	55%
NCI	34%	24%	84%

Figures 14–21 show the mean test-retest scatter plots and Bland-Altman LOA analysis for each eye and for each of the CCTs. Figures 22–24 summarize the Bland-Altman LOA for each eye and CCT so the LOA for each test can be more directly compared.

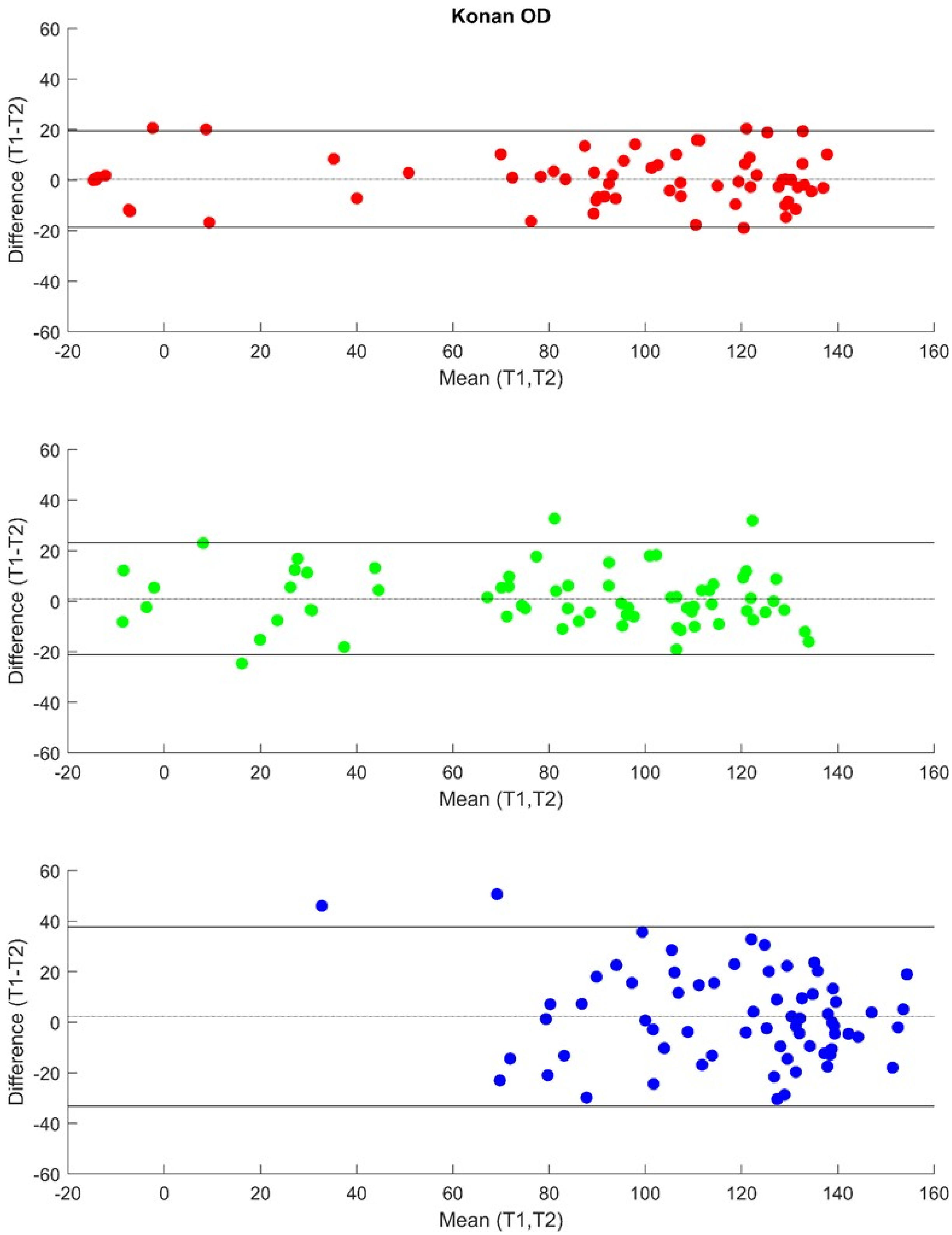

Figure 14. Konan OD Bland-Altman plots for L-cone (top), M-cone (center) and S-cone (bottom)

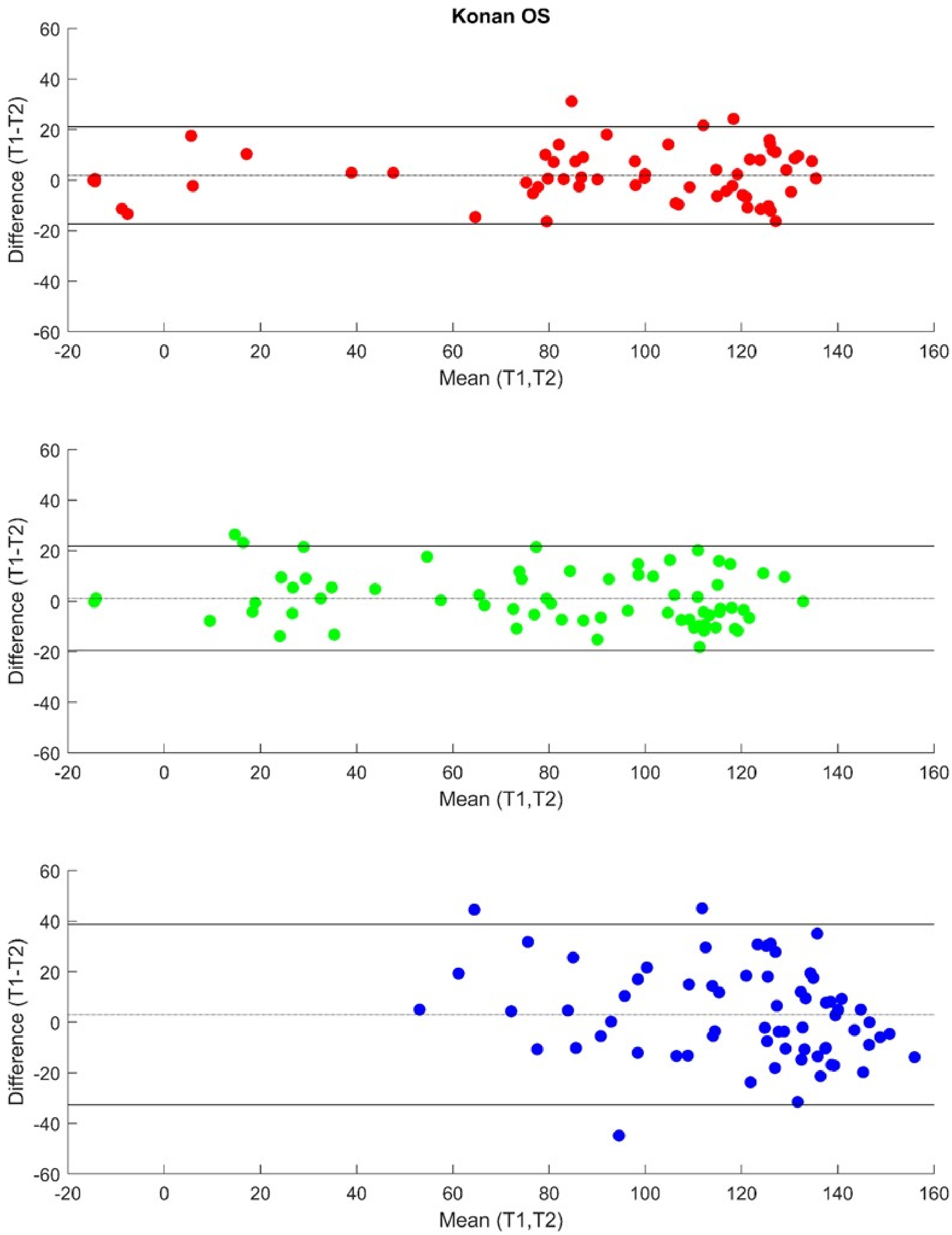

Figure 15. Konan OS Bland-Altman plots for L-cone (top), M-cone (center) and S-cone (bottom)

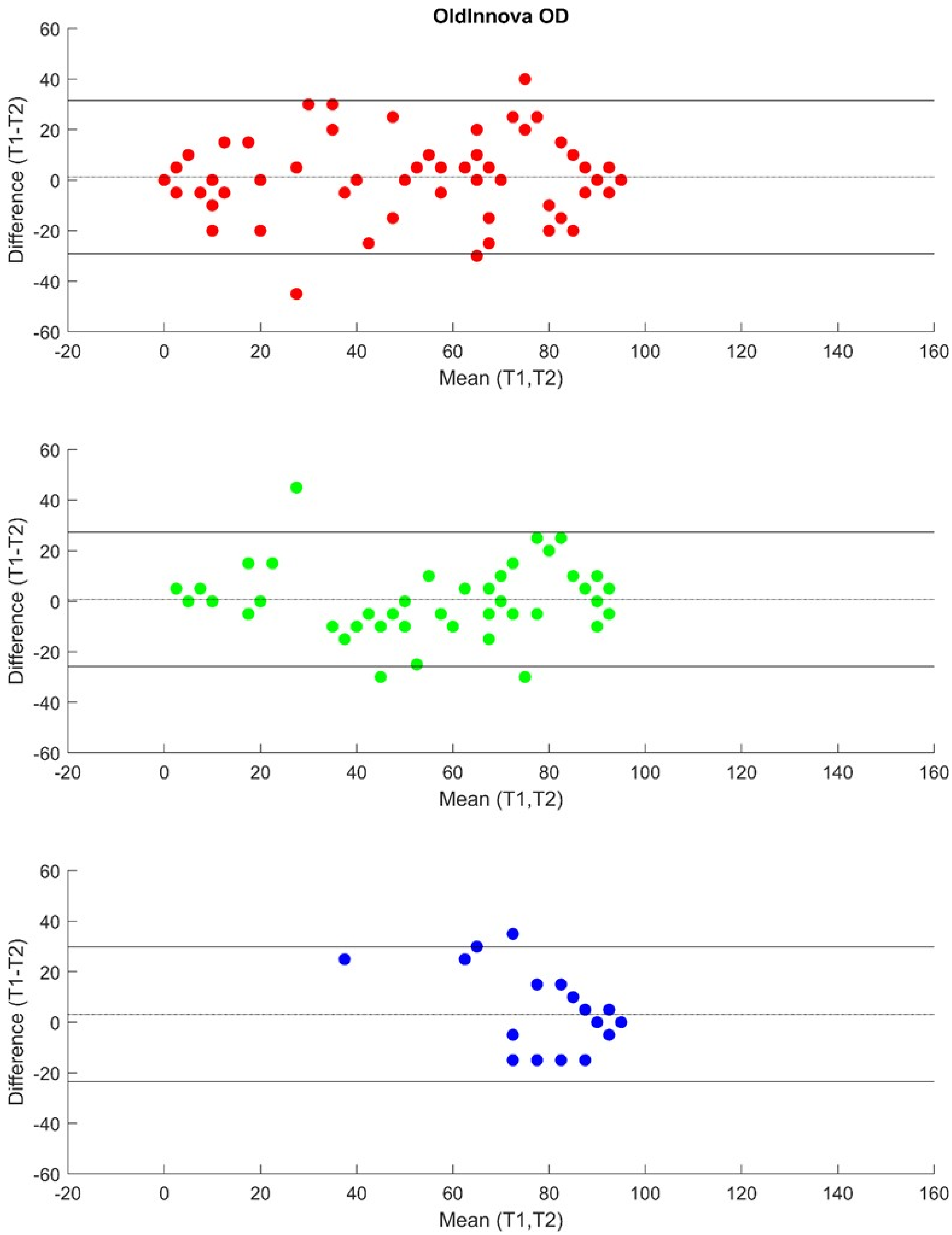

Figure 16. Old Innova OD Bland-Altman plots for L-cone (top), M-cone (center) and S-cone (bottom)

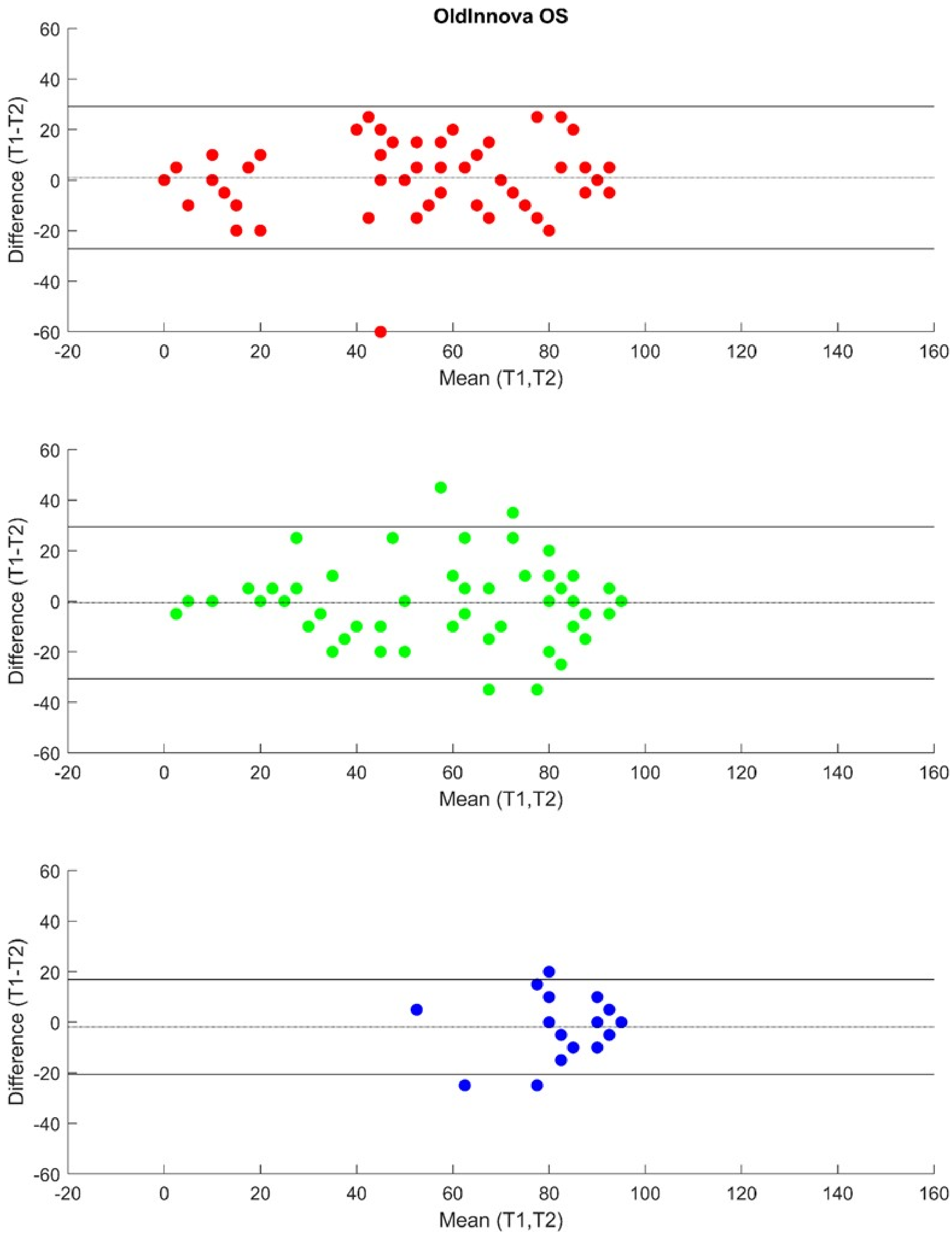

Figure 17. Old Innova OS Bland-Altman plots for L-cone (top), M-cone (center) and S-cone (bottom)

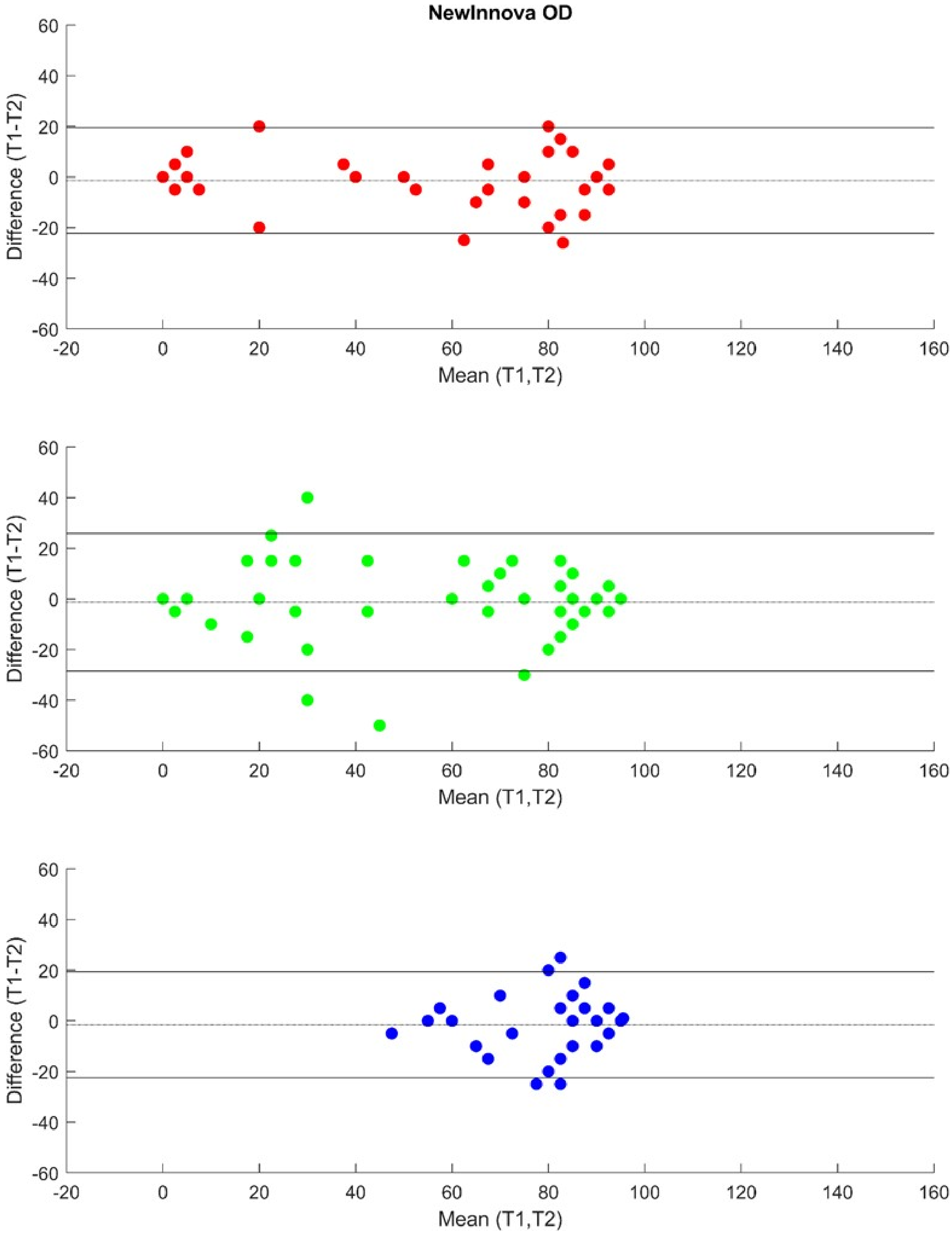

Figure 18. New Innova OD Bland-Altman plots for L-cone (top), M-cone (center) and S-cone (bottom)

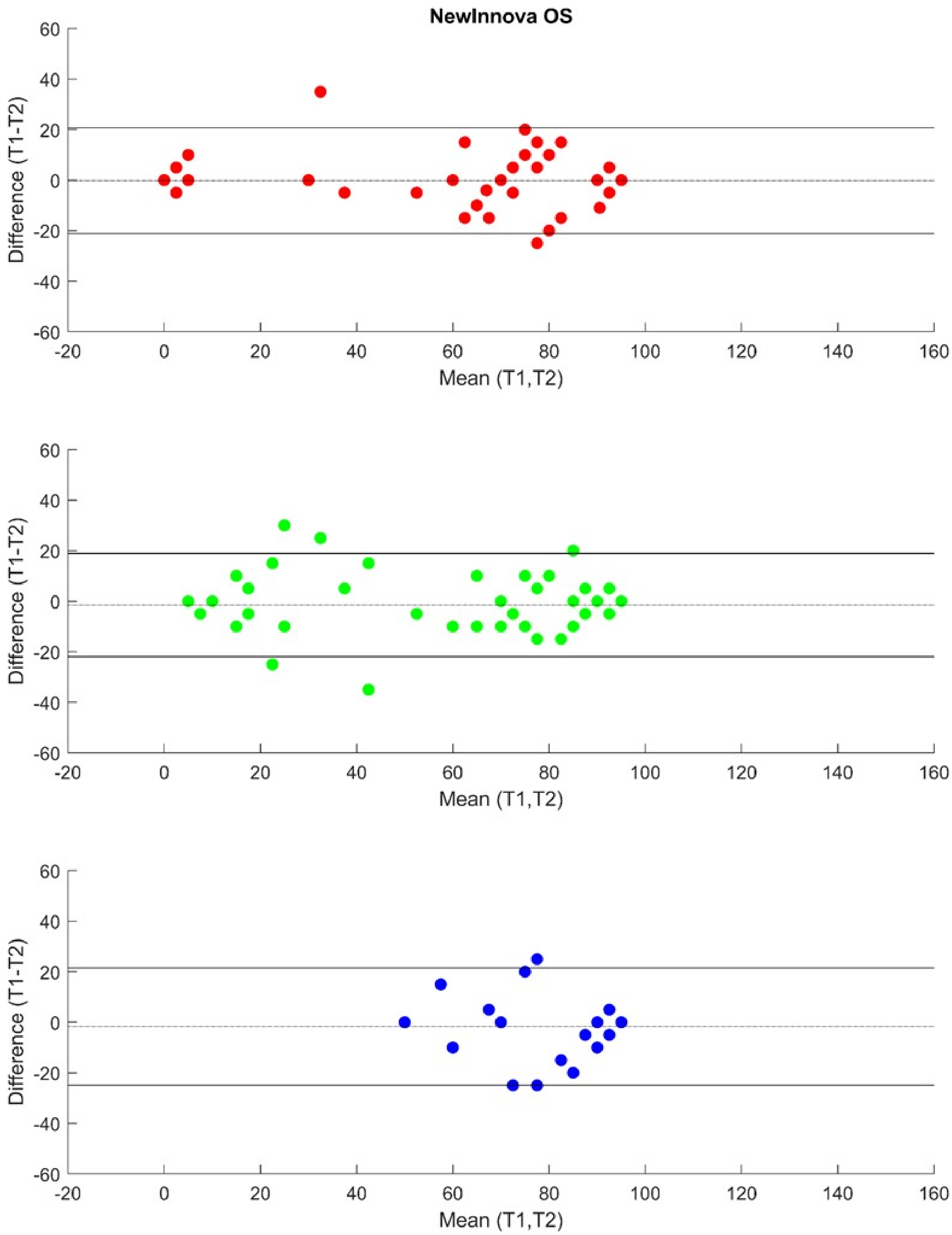

Figure 19. New Innova OS Bland-Altman plots for L-cone (top), M-cone (center), and S-cone (bottom)

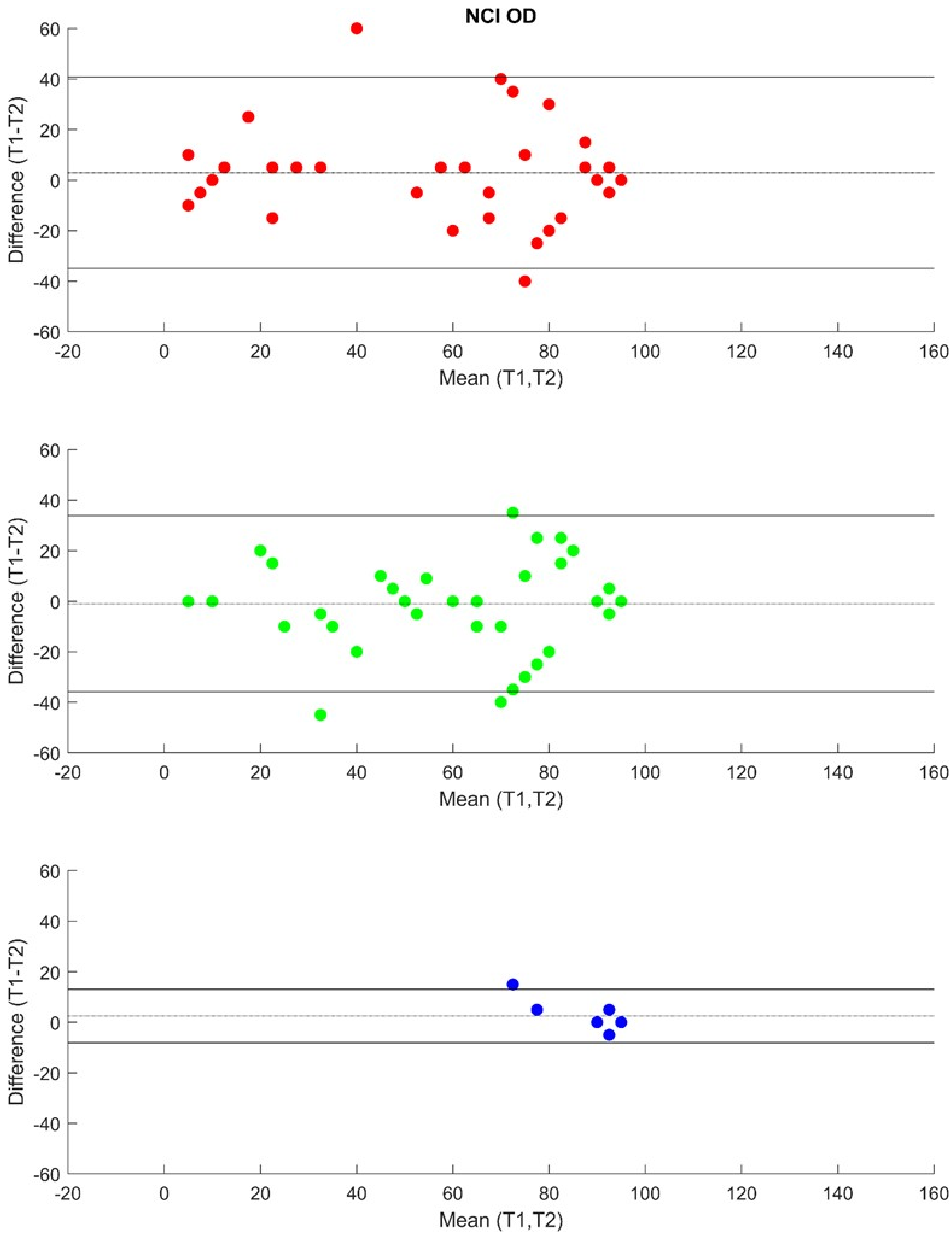

Figure 20. NCI OD Bland-Altman plots for L-cone (top), M-cone (center), and S-cone (bottom)

Distribution Statement A: Approved for public release. AFRL-2021-3506, cleared 26 October 2021

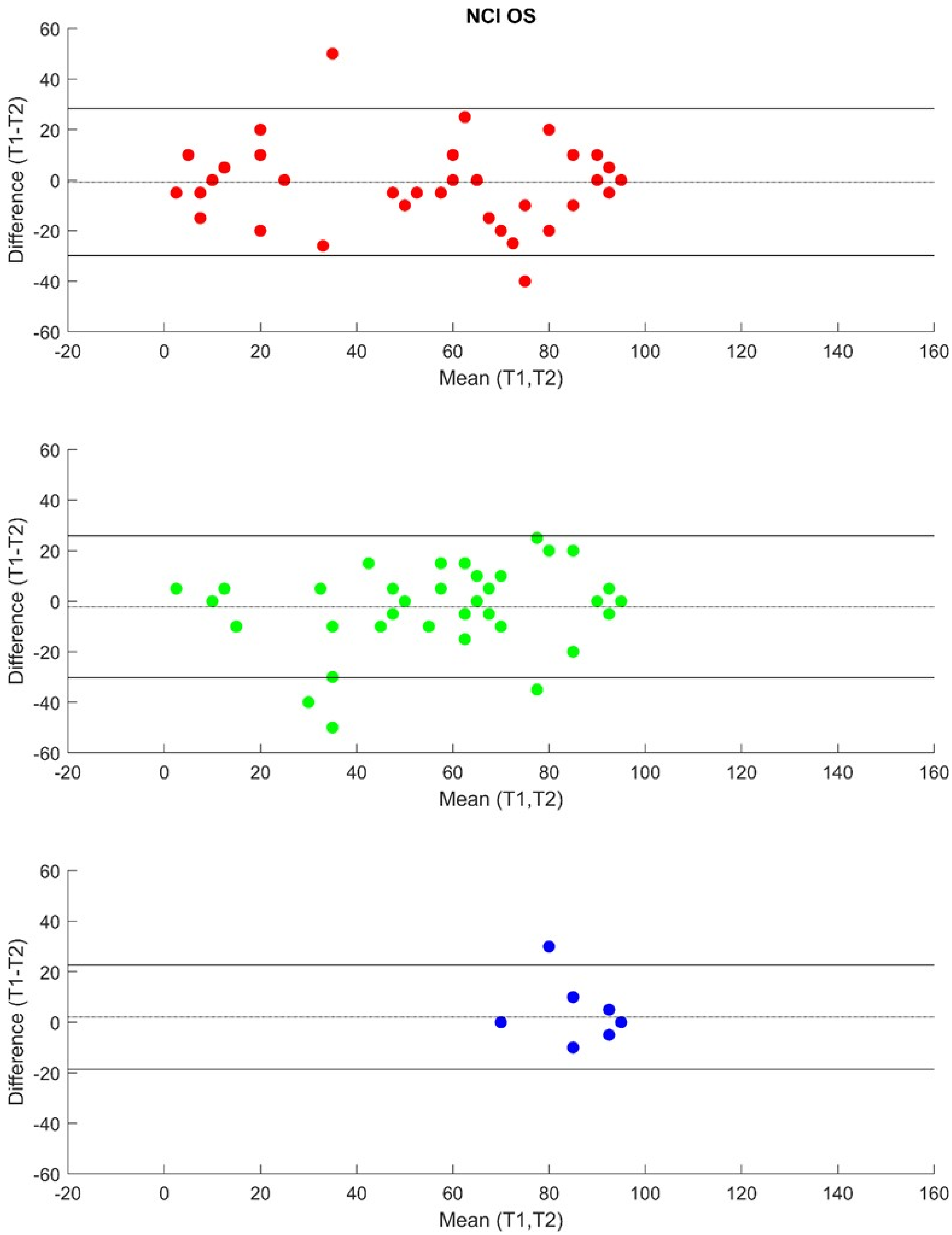

Figure 21. NCI OS Bland-Altman plots for L-cone (top), M-cone (center) and S-cone (bottom)

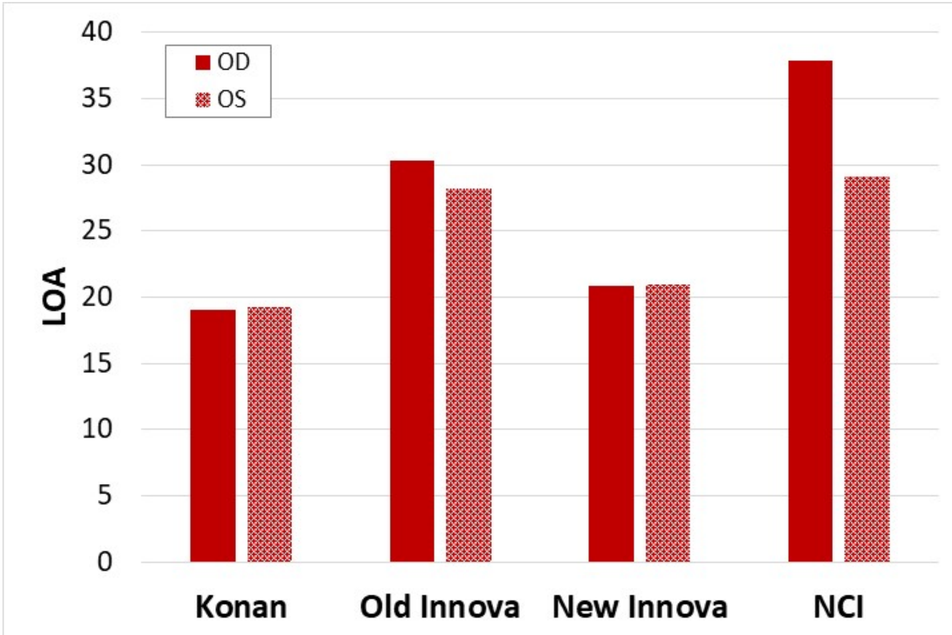

Figure 22. L-cone test-retest Bland-Altman limits of agreement (LOA) for each eye and each CCT

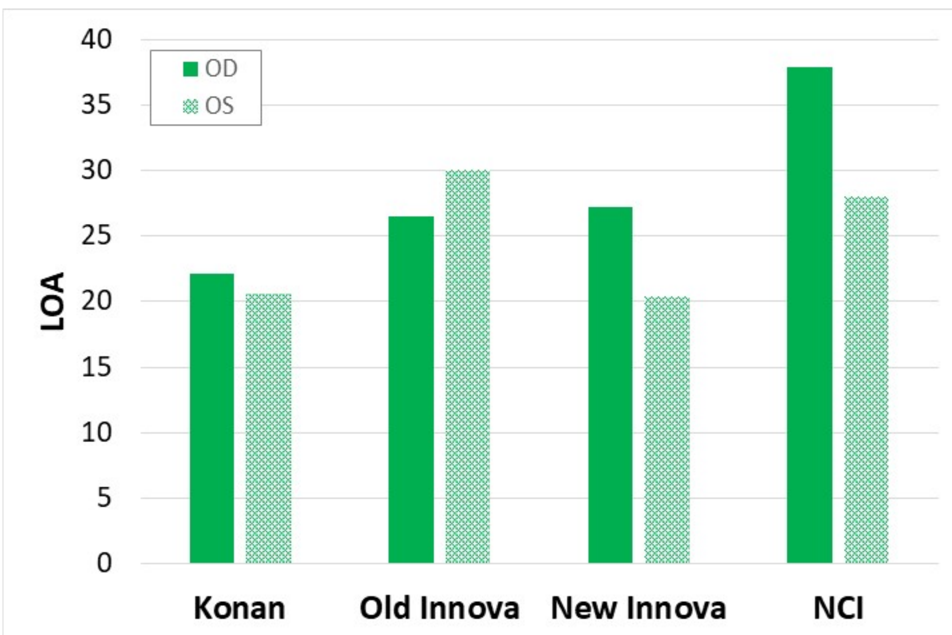

Figure 23. M-cone test-retest Bland-Altman limits of agreement (LOA) for each eye and each CCT

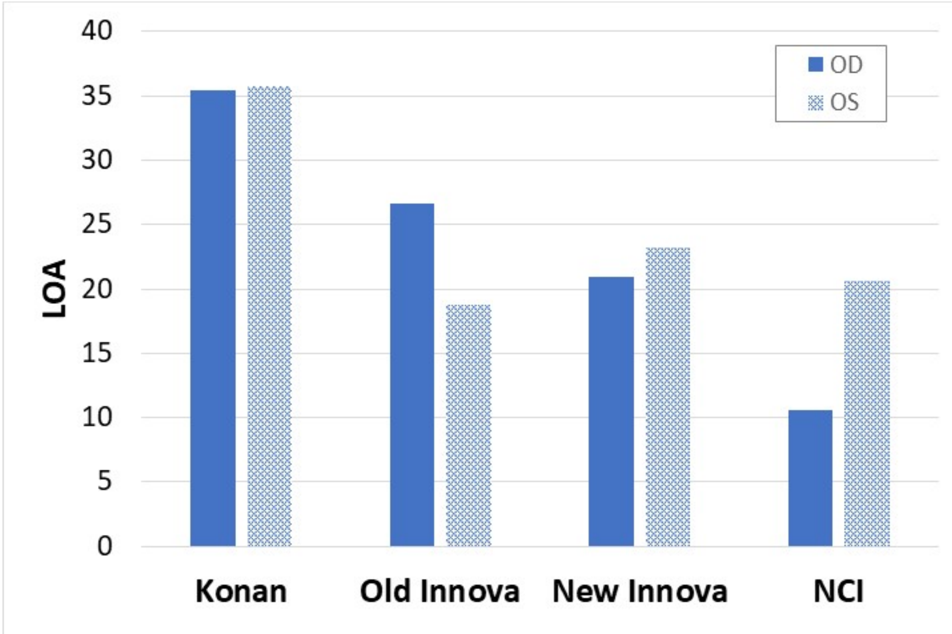

Figure 24. S-cone test-retest Bland-Altman limits of agreement (LOA) for each eye and each CCT

As shown, the LOA are lowest (i.e., best test-retest reliability) for the Konan CCT HD for the L- and M-cones. For the S-cone, the Konan device has the highest values. However, there was a very large ceiling effect for the S-cone test for both the Innova and NCI devices, with approximately 50% of subjects scoring at the maximum value of 100 for both of the Innova CCTs, and 84% of subjects scoring at the maximum value of 100 on the NCI CCT. Thus, it is not clear that this is a meaningful comparison. It is also notable that the NCI S-cone scores are compressed into a very narrow range of variability, which artificially limits the variance of the differences resulting in a lower LOA. There is also a surprisingly large difference between the OD and OS scores, particularly for the NCI. Interocular test scores would generally be expected to be very similar except for cases of disease or injury.

Figure 25 shows the test-retest scatter plots and R^2 value for each of the four tests for L-, M- and S-cones for the right eye (OD). This provides an alternative approach to examining the test-retest reliability. As shown, the Konan CCT HD has very good reliability for the L- and M-cone tests. Both the Bland-Altman and scatter plots reveal the substantial amount of quantization error that occurs with the Innova and NCI tests due to the use of only five contrast levels and their custom staircase method.

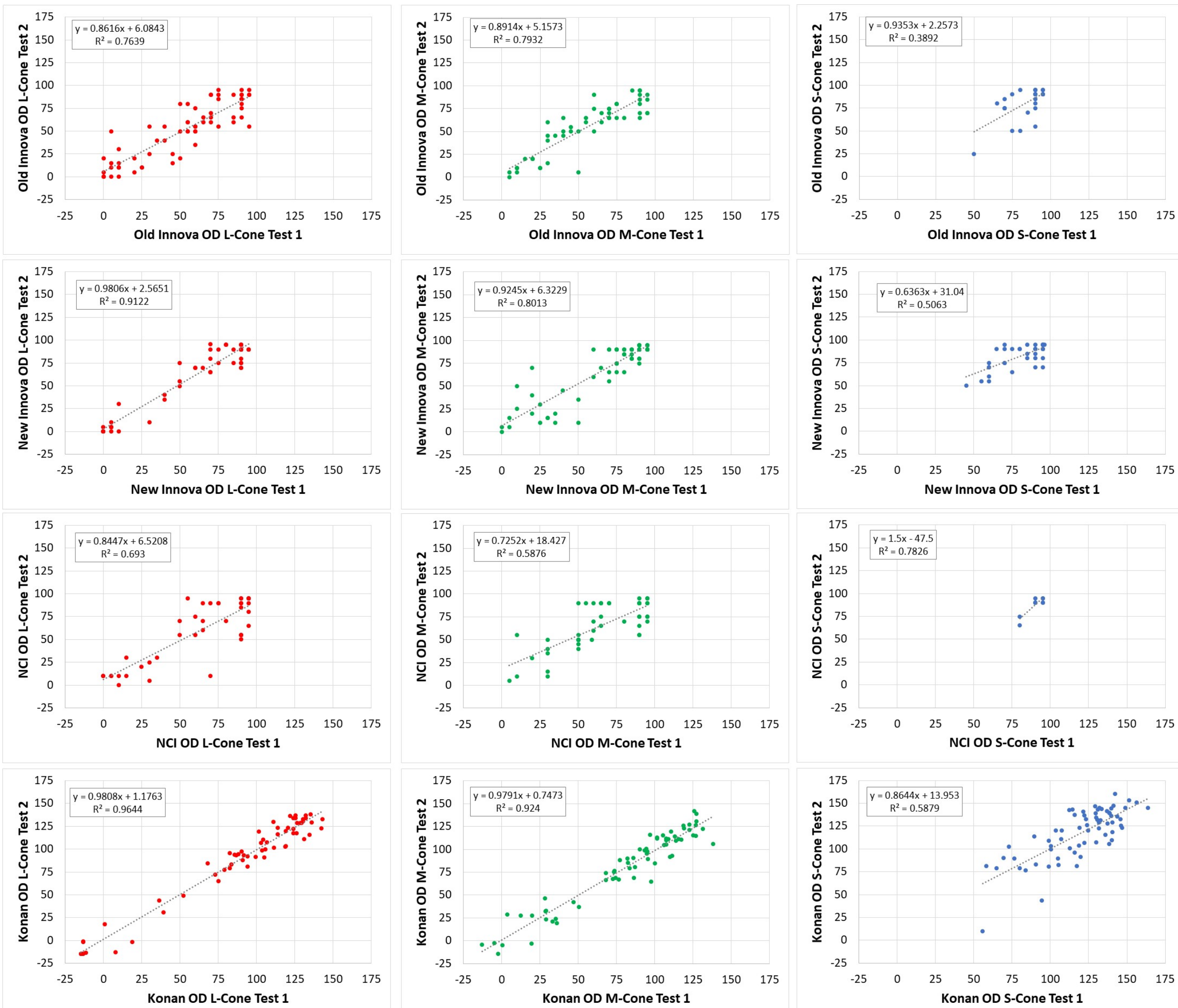

Figure 25. Test-retest scatter plots and R^2 value for each of the four tests for L-, M- and S-cones for the right eye (OD)

5.0 DISCUSSION

Although previous research has shown that the RCCT reliably screens for color deficiency (Rabin et al, 2011; Hovis & Almustanyir, 2017), this is the first study to compare all of the commercially available CCTs in terms of test-retest reliability.

It is important to note that, based on previous research (Winterbottom et al, 2017), a large ceiling effect is expected for the Innova and NCI CCTs. Figure 26 shows the distribution of monocular

CCT scores for the Innova version 12.1 (left) and the OBVA CCT (OCCT, right) from the previous study. As shown in Figure 26 (left), approximately 60% of the Airmen tested on the Innova CCT scored at the maximum value of 100.

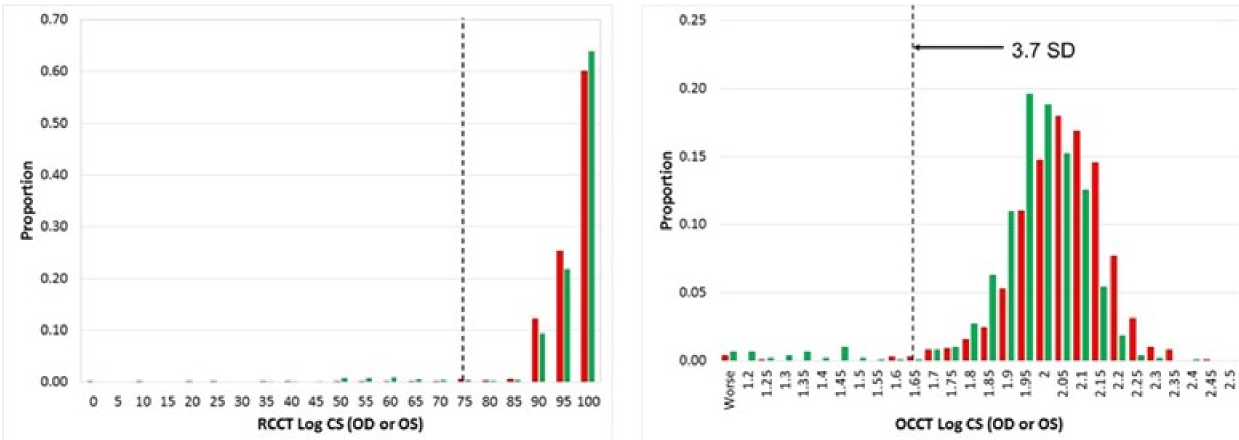

Figure 26. Left: Distribution of Innova RCCT scores (L-cone – 2,182 eyes tested; M-cone – 2,168 eyes tested). Right: Distribution of OCCT scores (L-cone – 1,091 eyes tested, M – cone – 1,084 eyes tested). Dashed line shows the USAF pass/fail criterion used until 2018.

However, it is notable that the proportion of participants scoring 100 differs substantially between versions of the Innova and NCI CCTs. This may be attributable to differing measured cone contrast values across the Innova and NCI CCTs. Figures 2–7 show the magnitude of the calibration error for each of the tests. Figure 27 shows the desired vs. measured contrast error while preserving the direction of the error for the L and M cones for the Innova and NCI CCTs. Although all three versions of the RCCT should use the same cone contrast values as shown in Table 1, the measured contrast reveals that optotype contrast can vary substantially.

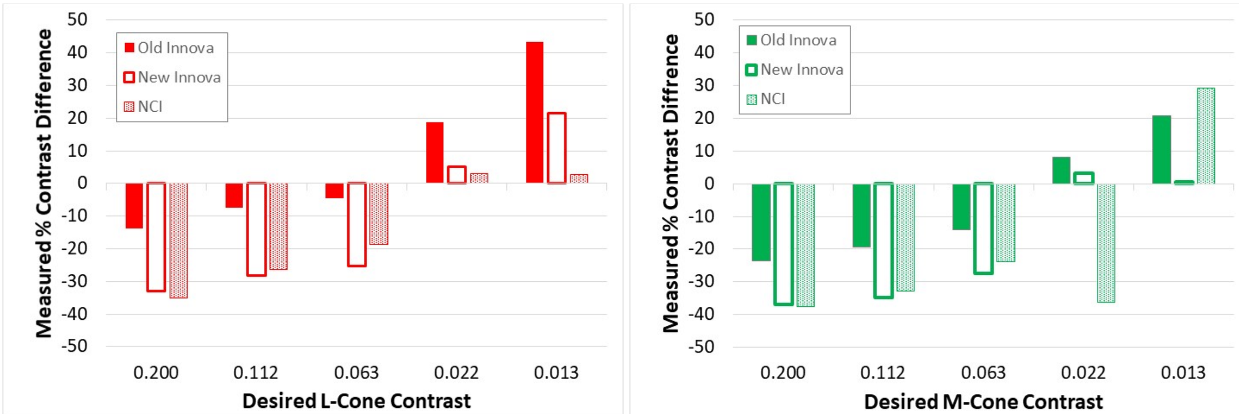

Figure 27 – Desired vs. measured cone contrast for L- and M-cone contrast at all available contrasts in the Old Innova, New Innova and NCI.

The OBVA Laboratory was established to support Air Force Surgeon General (AF/SG) objectives to modernize USAF vision standards and screening methods and supports the Human Performance Concept of Operations (CONOPS) published by the AF/SG in 2014. The objectives of the Human Performance CONOPS included:

- Define measurable health and human performance sustainment thresholds for all beneficiaries based on operational, occupational and/or personal, patient-oriented goals
- Centrally accumulate and analyze population data, determine effective evidence-based practices, and disseminate the knowledge to the healthcare teams caring for the representative populations
- Foster and institutionalize innovation and the dissemination of health and human performance sustainment knowledge throughout the enterprise
- Current approaches, depending on single point annually derived data sets fail to provide dynamic feedback on functionality given adverse intervening health impacts (for instance deployments, intervening health conditions). Essential to the CONOPS will be aggressive and deliberate development of required human performance measures, based on operator specific requirements.

More recently, the Air Force Medical Service (AFMS) has highlighted the need to develop vision screening metrics “capable of characterizing vision across the full performance spectrum using adaptive, threshold-level, visual performance metrics to support multiple USAF and Department of Defense (DoD) human performance optimization initiatives." AFMS leadership has also noted the importance of moving away from screening tests that indicate only coarsely presence/absence of disease.

Thus, it is essential that USAF vision screening tests go beyond simple pass/fail screening and are capable of characterizing the visual performance of individuals with normal, healthy vision, in addition to identifying potentially disqualifying visual dysfunction. Threshold-level vision screening tests with high test-retest reliability are needed to establish the quantitative relationships between an Airman's visual capability and operational performance to generate population data needed to optimize Airmen performance.

Good test-retest reliability is essential for research involving individual differences, and particularly for research intended to identify the relationship between individual differences and performance on other tasks. As Hedge, Powell, and Sumner (2018) discuss in their review of fundamental differences between experimental and correlational research, this is because the reliability of two measures will limit the magnitude of the correlation that can be observed between them. Spearman (1910) states that measures with poor reliability are ill-suited to correlational research, as the ability to detect relationships with other constructs will be compromised by the inability to effectively distinguish between individuals on that dimension.

Based on the research presented here, we conclude that the Konan CCT HD produces much more reliable color vision test results in comparison to other commercially available cone contrast tests and is therefore better suited for routine screening as well as research purposes. However, additional research is needed to examine the utility of the blue cone test, and users of any of the CCTs should be cautious about blue cone test results when making pass/fail decisions involving blue cone test scores.

6.0 REFERENCES

Barbur, J. L., & Rodriguez-Carmona, M. (2017). Colour vision requirements in visually demanding occupations. British medical bulletin, 122(1), 51-77.

Cole, B. L. (1993). Does defective colour vision really matter? In Colour vision deficiencies XI (pp. 67-86). Springer, Dordrecht.

Cole, B. L. (2004). The handicap of abnormal colour vision. Clinical and Experimental Optometry, 87(4-5), 258-275.

Steward, J. M., & Cole, B. L. (1989). What do color vision defectives say about everyday tasks?. Optometry and vision science: official publication of the American Academy of Optometry, 66(5), 288-295.

Delpero, W. T., O'Neill, H., Casson, E., & Hovis, J. (2005). Aviation-relevant epidemiology of color vision deficiency. Aviation, space, and environmental medicine, 76(2), 127-133.

Fee, J. (1879). Color-Blindness, and Railway Accidents. Transactions of the Kansas Academy of Science, 7, 29-37. doi:10.2307/3623576

Gaska, J., Winterbottom, M., and Van Atta, A. (2016). Operational Based Vision Assessment Cone Contrast Test: Description and Operation. Air Force Research Laboratory Technical Report AFRL-SA-WP-SR-2016-0007.

Gaska, J., Winterbottom, M., Hadley, S., O'Keefe, E., Shoda, E., and Van Atta, A. (2021). Evaluation of Konan Medical CCT HD. Air Force Research Laboratory Technical Report.

Giavarina, D. (2015). Understanding bland altman analysis. Biochemia medica, 25(2), 141-151.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. Behavior research methods, 50(3), 1166-1186.

Hovis, J. and Almustanyir, A. (2017). Assessment of the next generation of colour vision tests for pilots and aircrew. Defence Research and Development Canada Contractor Report (DRDC-RDDC-2017-C191).

Ishihara, S. (1918). Tests for color blindness. American Journal of Ophthalmology, 1(5), 376.

Prins, N. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. Frontiers in psychology, 9, 1250.

Rabin, J. (1996). Cone-specific measures of human color vision. Investigative ophthalmology & visual science, 37(13), 2771-2774.

Rabin, J. (2004). Quantification of color vision with cone contrast sensitivity. Visual neuroscience, 21(3), 483.

Rabin, J., Gooch, J., & Ivan, D. (2011). Rapid quantification of color vision: the cone contrast test. Investigative ophthalmology & visual science, 52(2), 816-820.

Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 1904-1920, 3(3), 271-295.

Stockman, A., & Sharpe, L. T. (2000). The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. Vision research, 40(13), 1711-1737.

Verriest, G. (1979). Vocational and practical implications of defective color vision. In Congenital and acquired color vision defects (pp. 349-357). Grune and Stratton New York.

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

%	Percent
AFMS	Air Force Medical Services
AF/SG	Air Force Surgeon General
CCT	Cone Contrast Test
C_{iso}	Contrast of the isolated cone type
CVD	Color Vision Deficient/Deficiency
CVN	Color Vision Normal
CONOPS	Human Performance Concept of Operations
L	Long wavelength
L-cone	Long wavelength (red) cone
LOA	Limits of Agreement
M-cone	Medium wavelength (green) cone
M	Middle wavelength
OBVA	Operational Based Vision Assessment
OCCT	OBVA Cone Contrast Test
OD	Oculus Dexter (right eye)
OS	Oculus Sinister (left eye)
PIP	Pseudoisochromatic Plate
RCCT	Rabin Cone Contrast Test
RGB	Red, Green, Blue
S	Short wavelength
S-cone	Short wavelength (blue) cone
T1	Test 1
T2	Test 2
USAF	United States Air Force