AFRL-RH-WP-TR-2021-0107

AIRCREW VISION STANDARDS RESEARCH AUTOMATED VISION TEST (AVT)

Eleanor M. O'Keefe, Elizabeth A. Shoda

KBR

Marc D. Winterbottom, James P. Gaska, Steven C. Hadley

711 HPW/RHBC

Dr. Mackenzie G. Glaholt

Defence Research and Development Canada

Dr. Maria Gavrilescu, Dr. Peter Gibbs

Defence Science and Technology Group

AUGUST 2021
Final Report

Distribution A: Approved for public release.

See additional restrictions described on inside pages

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-2021-0107 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signature//
MARC WINTERBOTTOM
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-YY)***
29-08-21

2. REPORT TYPE
Final

3. DATES COVERED ***(From - To)***
September 2018 - August 2021

4. TITLE AND SUBTITLE
Aircrew Vision Standards Research Automated Vision Test (AVT)

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

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
1Eleanor M. O'Keefe, Elizabeth A. Shoda
2Marc D. Winterbottom, James P. Gaska, Steven C. Hadley
3Dr. Mackenzie G. Glaholt
4Dr. Maria Gavrilescu, Dr. Peter Gibbs

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
Legacy RHM

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
1KBR, Dayton OH 45431
3Defence Research and Development Canada, Toronto, ON Canada
4Defence Science and Technology Group, Aerospace Division, Fishermans Bend, Victoria, Australia

8. PERFORMING ORGANIZATION REPORT NUMBER
TTCP Project HUM-TP22-2016-05.A.1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
2Air 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/MONITORING AGENCY ACRONYM(S)
711 HPW/RHBC

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RH-WP-TR-2021-0107

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

13. SUPPLEMENTARY NOTES
Report contains color. AFRL-2022-2033, Cleared 31 May 2022

14. ABSTRACT
Technology has changed dramatically since vision screening tests currently used by the United States Air Force (USAF), Royal Canadian Air Force (RCAF), and Royal Australian Air Force (RAAF) were adopted post-WWII, and vision screening standards and technologies must be updated accordingly. A collaborative research team from three nations was established through The Technical Cooperation Program (TTCP) to pursue research concerning aircrew vision standards. This report summarizes research with the following objectives: 1) Evaluate test-retest reliability of each test and feasibility of each test for use as a standard screening test; 2) Compare results for two versions of the Automated Vision Test (AVT) – an older version and an updated version; 3) Examine the consistency of the AVT results at different labs and with different populations; 4) Compare results for two alternative stereo acuity tests; 5) Develop a normative database for each of the AVT tests to support development of pass/fail criteria for a new set of modernized, computer-based vision tests for military vision screening, and 6) Identify candidate tests suitable for replacing selected existing tests and refine test methods and instructions. The AVT described here provides the basis for a new generation of more accurate and reliable, computer-based vision tests to support earlier diagnosis of eye disease, more accurate monitoring of treatment and recovery, and occupational vision screening.

15. SUBJECT TERMS
Automated Vision Test (AVT), vision standards, vision screening, stereo acuity, cone contrast test

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			Marc Winterbottom
Unclassified	Unclassified	Unclassified	SAR	71	19b. TELEPHONE NUMBER *(Include Area Code)* N/A

The
Technical
Cooperation
Program

Australia - Canada - New Zealand - United Kingdom - United States of America

---

TTCP TECHNICAL REPORT

AIRCREW VISION STANDARDS RESEARCH AUTOMATED VISION TEST

29 August 2021

TTCP Project HUM-TP22-2016-05.A.1

TABLE OF CONTENTS

Page

LIST OF FIGURES ... iii
LIST OF TABLES ... iv
PREFACE ... v
1.0 SUMMARY ... 1
2.0 BACKGROUND ... 1
3.0 METHODS ... 3
3.1 Participants ... 3
3.2 Apparatus ... 3
3.3 Vision Measures ... 4
3.3.1. Landolt C Contrast Sensitivity Test (LCST) ... 4
3.3.2. OBVA Cone Contrast Test (OCCT) ... 5
3.3.3. Dual Ring Stereo Acuity Test ... 5
3.3.4. Stereo Search Test ... 6
3.3.5. Fusion Range Test ... 7
3.3.6. OBVA Motion Coherence Test 2.0 ... 8
3.4 Design and Procedures ... 9
3.4.1. Comparing the old and new AVT ... 9
3.5 Statistical Analysis ... 12
4.0 RESULTS ... 12
4.1 Study Sample ... 12
4.2 Descriptive Statistics ... 12
4.3 Test-Retest Reliability, Differences between Versions and Labs ... 15
4.3.1. LCST ... 15
4.3.2. OCCT ... 19
4.3.3. Dual Ring Stereo Acuity ... 28
4.3.3.1. Dual Ring Stereo Acuity Test – Effect of Training and Instructions ... 33
4.3.4. Stereo Search ... 34
4.3.5. Fusion Range ... 34
4.3.6. Motion Coherence ... 37
5.0 DISCUSSION ... 38
5.1 Recommendations for Future Work ... 40

6.0 CONCLUSION 41
7.0 REFERENCES 43
LIST OF ACRONYMS 46
Appendix A – Instructions for each test included in the AVT Battery 47
Appendix B – AVT Histograms 52

LIST OF FIGURES

Page

Figure 1. Logitech game controller used to enter responses for vision tests. 4

Figure 2. An example of the size reticle (left). Landolt C with the gap positioned on the left and relatively high contrast (center). Landolt C with upward orientation and low contrast (right). 5

Figure 3. OCCT L, M, and S Landolt C optotypes. 5

Figure 4. Dual ring stereo acuity test stimulus. 6

Figure 5. AVT SST stimulus 7

Figure 6. Horizontal fusion range stimuli (left and right eye images). 8

Figure 7. Motion coherence test stimuli. 9

Figure 8. Instruction and demo screen for the near dual ring test in the new AVT battery 11

Figure 9. Contrast Sensitivity test-retest scores for OBVA data for contrast sensitivity with a Landolt C stimulus with 1.25, 2.5, and 16.7 arcmin gap sizes. Left: old AVT. Right: new AVT 16

Figure 10. Contrast sensitivity Bland-Altman graphs for contrast sensitivity with a Landolt C stimulus with 1.25, 2.5, and 16.7 arcmin gap sizes. Left: old AVT. Right: new AVT. 17

Figure 11. Acuity test-retest scores. Left: old AVT. Right: new AVT. 18

Figure 12. Acuity Bland-Altman graphs. Left: old AVT. Right: new AVT. 19

Figure 13. OCCT L cone test-retest scores for OBVA old AVT, OBVA new AVT, and DRDC. 20

Figure 14. OCCT M cone test-retest scores for OBVA old AVT, OBVA new AVT, and DRDC old AVT. 21

Figure 15. OCCT S cone test-retest scores for OBVA old AVT, OBVA new AVT, and DRDC old AVT. 22

Figure 16. OCCT L cone Bland Altman graphs for OBVA old AVT, OBVA new AVT, and DRDC old AVT. 23

Figure 17. OCCT M cone Bland Altman graphs for OBVA old AVT, OBVA new AVT, and DRDC old AVT. 24

Figure 18. OCCT S cone Bland Altman graphs for OBVA old AVT, OBVA new AVT, and DRDC. 25

Figure 19. Histogram showing spread of binocular OCCT L cone (red) and M cone (green) scores. Cutoff score for color vision abnormals is 3.7 standard deviations from the mean. From Winterbottom, Gaska, Wright, & Hadley, 2017. International Color Vision Society 24th Symposium. 26

Figure 20. OCCT probability density graphs for OBVA old AVT, OBVA new AVT, and DRDC. 27

Figure 21. Dual Ring Far test-retest scores for OBVA old AVT, OBVA new AVT, and DSTG old AVT. 29

Figure 22. Dual Ring Near test-retest scores for OBVA old AVT, OBVA new AVT, and DSTG old AVT. ... 30

Figure 23. Dual Ring Far Bland Altman graphs for OBVA old AVT, OBVA new AVT, and DSTG old AVT. ... 31

Figure 24. Dual Ring Near Bland Altman graphs for OBVA old AVT, OBVA new AVT, and DSTG old AVT. ... 32

Figure 25. Dual Ring Far and Near boxplots for OBVA old AVT, OBVA new AVT, and DSTG old AVT. ... 33

Figure 26. Average near dual ring stereo acuity for subjects at three different Army bases in Australia. Dual ring test instructions and training were refined for each successive location. ... 34

Figure 27. Left: Stereo search test-retest scores. Right: Stereo search Bland-Altman graph ... 34

Figure 28. Horizontal fusion range test-retest scores for OBVA old AVT and new AVT ... 35

Figure 29. Vertical fusion range test-retest scores for OBVA old AVT and new AVT ... 36

Figure 30. Horizontal fusion range Bland-Altman graphs for OBVA old AVT and new AVT.. 36

Figure 31. Vertical fusion range Bland-Altman graphs for OBVA old AVT and new AVT ... 37

Figure 32. Motion test-retest scores. ... 38

Figure 33. Motion Test Bland-Altman graphs ... 38

LIST OF TABLES

Page

Table 1. AVT Time Settings. ... 10

Table 2. Descriptive statistics for each of the AVT tests. ... 13

Table 3. Descriptive statistics for each of the AVT tests (continued). ... 14

Table 4. Pearson correlation statistics for the contrast sensitivity LCST tests. ... 15

Table 5. Pearson correlation statistics for the 4M Acuity test. ... 18

Table 6. Pearson correlation statistics for the OCCT tests. ... 19

Table 7. Descriptive statistics for the OCCT tests for color normals only. ... 26

Table 8. Pearson correlation statistics for old and new dual ring stereo acuity tests. ... 28

Table 9. Pearson correlation statistics for the fusion range tests. ... 35

Table 10. Pearson correlation statistics for the motion coherence tests. ... 37

Table 11. Differences between mean test scores. ... 40

PREFACE

The views presented herein are those of the authors and do not reflect views of the United States Air Force, the Department of Defense, or any other government agency. All studies described within this report were accomplished in accordance with Air Force Research Laboratory Institutional Review Board-approved research protocols. The DRDC Toronto Research Centre data collection protocols (#2014-044 and #2017-009) were reviewed and approved by the DRDC Toronto Research Centre Human Research Ethics Committee. For DSTG, ethics was approved by the Chief of Aerospace Division and was registered with the DSTG Low Risk Ethics Panel (AD 01-17). All participants provided informed consent prior to participation.

1.0 SUMMARY

Technology has changed dramatically since most of the vision screening tests currently used by the United States Air Force (USAF), Royal Canadian Air Force (RCAF), and Royal Australian Air Force (RAAF) were first adopted post-WWII, and vision screening standards and technologies must be updated accordingly. A collaborative research team from three nations was established through The Technical Cooperation Program (TTCP) to pursue research concerning aircrew vision standards. Through additional partnering and contracting arrangements, resources and expertise at York University in Toronto, ON, Canada, the University of Waterloo, in Waterloo, ON, Canada, the University of Melbourne, Melbourne, VIC, Australia, and Deakin University, Geelong, VIC, Australia were also engaged to support project objectives. This report summarizes research with the following objectives: 1) Evaluate test-retest reliability of each test and feasibility of each test for use as a standard screening test; 2) Compare results for two versions of the AVT – an older version and an updated version; 3) Examine the consistency of the AVT results at different labs and with different populations; 4) Compare results for two alternative stereo acuity tests (dual ring and stereo search test); 5) Develop a normative database for each of the AVT tests to support development of pass/fail criteria for a new set of modernized, computer-based vision tests for military vision screening, and 6) Identify candidate tests suitable for replacing selected existing tests and refine test methods and instructions. The AVT software developed by 711 HPW and shared with the TTCP partner nations proved to be an enabling technology that supported research collaboration at 9 different laboratories/clinics in Australia, Canada, and the United States. The AVT described here provides the basis for a new generation of more accurate and reliable, commercially available, computer-based vision tests to support earlier diagnosis of eye disease, more accurate monitoring of treatment and recovery, and occupational vision screening.

2.0 BACKGROUND

Technology has changed dramatically since most of the vision screening tests currently used by the USAF, RCAF, and RAAF were first adopted post-WWII, and vision screening standards and technologies must be updated accordingly. Vision screening methods developed to screen aircrew to view distant targets with unaided vision may not be adequate for aircrew using advanced color-coded display technology and vision enhancement devices such as night vision devices and thermal imagers, helmet-mounted displays, hyper-stereoscopic remote vision systems, and many other advanced technologies now routinely used by military personnel. Furthermore, new research is demonstrating that injury/disease states can be identified earlier, and therefore treated sooner, through the use of more accurate and reliable tests. Additionally, treatments for injury/disease can be tracked more carefully with the use of more accurate/repeatable tests, which could allow faster return to service for military personnel, improving readiness. Similarly, the effects of degraded environments on vision, such as hypoxic conditions, can be more carefully evaluated using more accurate and repeatable vision tests. Finally, the vision screening requirements for aircrew requires substantial resources, which often include travel and continued monitoring throughout the careers of many military personnel. Increasing the accuracy, repeatability, and efficiency of these tests is expected to result in substantial cost savings and improved health monitoring.

A collaborative research team from three nations was established through The Technical Cooperation Program (TTCP) to pursue research concerning aircrew vision standards. Through additional partnering and contracting arrangements, resources and expertise at York University in Toronto, ON, Canada, the University of Waterloo, in Waterloo, ON, Canada, the University of Melbourne, Melbourne, VIC, Australia, and Deakin University, Geelong, VIC, Australia were also engaged to support project objectives. The objectives of this project were to 1) To develop and validate computer-based tests for color and depth perception for aircrew; 2) to research the relationship between vision test metrics and operational performance; 3) to modernize vision standards for military aviation personnel; 4) to support interoperability and joint training through more closely aligned medical standards.

Towards these objectives, the Air Force Research Laboratory's Operational Based Vision Assessment (OBVA) Laboratory developed a battery of computer-based Automated Vision Tests (AVT). In September 2018, a new software update was made to the AVT battery. The update included changes to participant instructions, stimulus presentation times, on-screen examples, and the addition of a new stereo acuity test. This report summarizes AVT data collected through August 22, 2019 and compares the results of the "old AVT" (data collected using software prior to version 1.2) to the "new AVT" (software versions 1.2 and 1.3.7), including descriptive statistics for each test as well as test-retest reliability. Additionally, old AVT data collected at Defence Research and Development Canada (DRDC) and at Defence Science and Technology Group (DSTG) in Australia were compared to data collected by the OBVA Lab. The objectives of the research described in this report were:

1. Evaluate test-retest reliability of each test and feasibility of each test for use as a standard screening test.
2. Compare results for two versions of the AVT – an older version and an updated version.
3. Examine the consistency of the AVT results at different labs and with different populations.
4. Compare results for two alternative stereo acuity tests (dual ring and stereo search test).
5. Develop a normative database for each of the AVT tests to support development of pass/fail criteria for a new set of modernized, computer-based vision tests for military vision screening.
6. Identify candidate tests suitable for replacing selected existing tests and refine test methods and instructions.

These data were collected in support of The Technical Cooperation Project (TTCP) international collaborative research agreement between the United States Air Force, Air Force Research Laboratory, 711th Human Performance Wing, DRDC, and DSTG. This research collaboration ("2016-05.A.1 Aircrew Vision Standards Research") was approved by the TTCP National Leads in November 2016 and was carried out under the HUM Group, Military Medicine Panel (TP22).

This report focuses on AVT data collection and evaluation. Additional reports and presentations were completed as part of this collaborative effort that compared the AVT OBVA cone contrast test (OCCT) to other color screening tests (Hovis, Almustanyir, & Glaholt, 2019; Hovis, Almustanyir, & Glaholt, 2018; Hovis, Almustanyir, & Glaholt, 2017; Hovis & Almustanyir, 2017) and evaluated the relationship between vision tests and operationally-relevant performance (Gaska, Winterbottom, Fu, Eisenhauer, & Hadley, 2019; Hartle, Sudhama, Deas, Allison, Irving, Glaholt, & Wilcox, 2019a, 2019b; Sudhama, Hartle, Allison, Wilcox, & Irving, 2019; Abel, Karas, Douglass, Gavrilescu, & Gibbs, 2019; Gavrilescu, Karas, Douglass, Abel, & Gibbs, 2019;

Douglass, Karas, Gavrilescu, White, Gibbs, & Abel, 2018; Karas, Gavrilescu, Douglass, White, Gibbs, Abel, Winterbottom, & Hadley, 2018; O'Keefe, Gaska, Bullock, Winterbottom, & Hadley, 2018; Winterbottom, Lloyd, Gaska, Williams, Wright, & Hadley, 2017; Gaska, Winterbottom, Bullock, & Hadley, 2017; Deas, Allison, Hartle, Irving, Glaholt, & Wilcox, 2017).

3.0 METHODS

3.1 Participants

This study was accomplished in accordance with Air Force Research Laboratory Institutional Review Board protocol number FWR20170095H. This study sought to recruit as many participants as practical to collect a large set of normative data for each AVT test with no exclusion criteria. For the OBVA Laboratory, participants were recruited from the Dayton, Ohio area. While color abnormals were not recruited specifically for the AVT battery, they were recruited for other testing being done in the lab and went through AVT testing as well. The DRDC Toronto Research Centre data collection protocols (#2014-044 and #2017-009) were reviewed and approved by the DRDC Toronto Research Centre Human Research Ethics Committee. For DRDC, participants in the color tests included undergraduate students from the University of Waterloo in Ontario, Canada as well as local residents. Here, DRDC used an anomaloscope to recruit a similar number of color normal and color deficient participants. Participants in the stereo acuity tests were recruited from the Department of National Defence and included soldiers and some civilian research staff. For DSTG data collection N=72 Australian Army Aviation (pilots and aircrew) were recruited across three different bases in Australia. The acquisition protocol was refined and improved across the three bases (improved instructions, practice, and increased stimulus presentation time from 2 to 4 sec) resulting in markedly different distributions of results. The final acquisition protocol was then used for an additional group of N=42 participant recruited from DSTG staff as well as staff and students from the University of Melbourne. To prevent any bias in interpretation of results due to changes in instructions, only the last group of participants was included in the comparison across OBVA Laboratory, DRDC and DSTG. Ethics was approved by the Chief of Aerospace Division and was registered with the DSTG Low Risk Ethics Panel (AD 01-17). All participants provided informed consent prior to participation.

3.2 Apparatus

All stimuli at each lab were generated using similar equipment. Standard Windows-based PCs with Nvidia graphics cards were used. The Landolt C Contrast Sensitivity Test (LCST) and the OBVA Cone Contrast Test (OCCT) were displayed on an NEC MultiSync P232W monitor with 1920 x 1080 pixels. The dual ring tests, the stereo search test, and the fusion range tests were displayed on Asus VG278HE or BenQ XL2411Z 3D monitors with 1920 x 1080 pixels that were compatible with a Nvidia 3D Vision2 kit using active shutter glasses. The motion perception tests were also displayed on the Asus monitor, but were not in 3D. Participants used the color-coded buttons on a Logitech game controller (Figure 1) to enter responses for all tests, except the motion perception tests where participants used the arrows keys on a standard computer keyboard.

Figure 1. Logitech game controller used to enter responses for vision tests. Photo by OBVA Lab personnel.

3.3 Vision Measures

3.3.1. Landolt C Contrast Sensitivity Test (LCST)

The LCST included both contrast sensitivity and visual acuity tests. During the LCST, a Landolt C was presented on the display with the gap in the C at four possible positions: left, right, top, or bottom. The participant's task was to identify the gap location. Across trials, the Landolt C varied in contrast as shown in Figure 2. During the visual acuity test a high contrast Landolt C varied in size. As shown in Figure 2, a reticle appeared prior to the presentation of the optotype to indicate the next trial and signal the size of the optotype. For contrast sensitivity testing, the Landolt C set had three sizes: 83.33, 12.5, and 6.25 arc minutes (16.7, 2.5, 1.25 arcmin gap sizes). The contrast of the Landolt C was varied according to the Psi adaptive procedure (Prins, 2018) for the contrast test. For acuity testing, the size of the Landolt C was varied according to the adaptive procedure. The slope of the psychometric function remained variable for each test. Going forward, the slope could be set from the average of this large data set. Both tests were administered at a four-meter distance. This test yielded the following measures: contrast sensitivity with a 1.25 arcmin gap Landolt C (*CS 1.25*), contrast sensitivity with a 2.5 arcmin gap Landolt C (*CS 2.5*), and contrast sensitivity with a 16.7 arcmin gap Landolt C (*CS 16.7*), all reported in log threshold contrast. The acuity test given at four meters yielded the measure *4M Acuity* reported in log Minimum Angle of Resolution (MAR). All four measures were produced in both the old and new AVT.

Figure 2. An example of the size reticle (left). Landolt C with the gap positioned on the left and relatively high contrast (center). Landolt C with upward orientation and low contrast (right).

3.3.2. OBVA Cone Contrast Test (OCCT)

The OCCT was similar to the LCST except that, instead of varying luminance contrast, the L, M, and S cone (i.e., long [red], medium [green], and short [blue] wavelength receptor) contrast was varied according to the adaptive procedure (see Gaska et al., 2016 for a detailed description of the OCCT). The colors red, green, and blue were selected to isolate the three cones for individuals with normal color vision. The size of the L, M, and S Landolt C's were 83.33 arc minutes. Figure 3 illustrates the appearance of the L, M, and S Landolt C optotypes. Like the LCST, the participant's task was to identify the gap location of the C. In the OBVA lab, the test was administered at both a one-meter and four-meter distance (see 3.4.1). At DRDC, the viewing distance was set at one meter. This test was given binocularly and yielded the following measures: *L cone, M cone,* and *S cone* reported in log threshold contrast for the old and new AVT. The slope of the psychometric function remained variable for each cone contrast test.

Figure 3. OCCT L, M, and S Landolt C optotypes.

3.3.3. Dual Ring Stereo Acuity Test

The dual ring stereo acuity test required observers to discriminate between crossed versus uncrossed disparity (i.e., decide whether the inner circle appeared in front of or behind the reference plane), and so required only a simple in-front-of or behind response. The disparity of the rings was varied according to the Psi adaptive procedure. The slope of the psychometric function remained variable for this test. Figure 4 shows the stimulus used for the dual ring stereo acuity test, which is similar in appearance to the Armed Forces Vision Tester (AFVT) circles test or the AO Vectograph circles test. The test was administered at two distances: 1 m (near) and 4 m (far). The software scaled the size of the ring stimuli according to viewing distance. The test was designed to minimize monocular cues and used antialiasing/blurring to support subpixel shifts of the position of each circle, which enabled very small changes in disparity. This test yielded the following measures: *Dual Ring Far* and *Dual Ring Near* reported in log arcseconds for the old and new AVT.

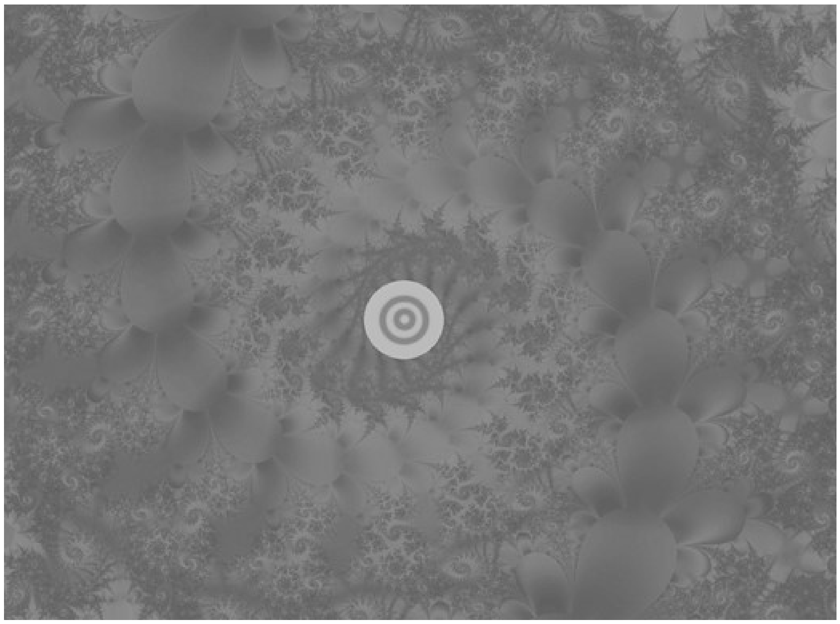


Figure 4. Dual ring stereo acuity test stimulus.

3.3.4. Stereo Search Test

The stereo search test (SST) also measured stereo acuity thresholds, wherein four rings were presented in a circle (top, bottom, left, right; see Figure 5). Like the dual ring stereo acuity test, monocular cues were minimized and subpixel shift was implemented. In each trial, one of the four rings appeared in front of the screen, one appeared behind the screen, and two were in-plane with the screen. Participants indicated which circle appeared in front of the screen. This test was administered at one meter only. This test yielded the following measure: *Stereo Search* reported in log arcseconds. The slope of the psychometric function remained variable for this test.

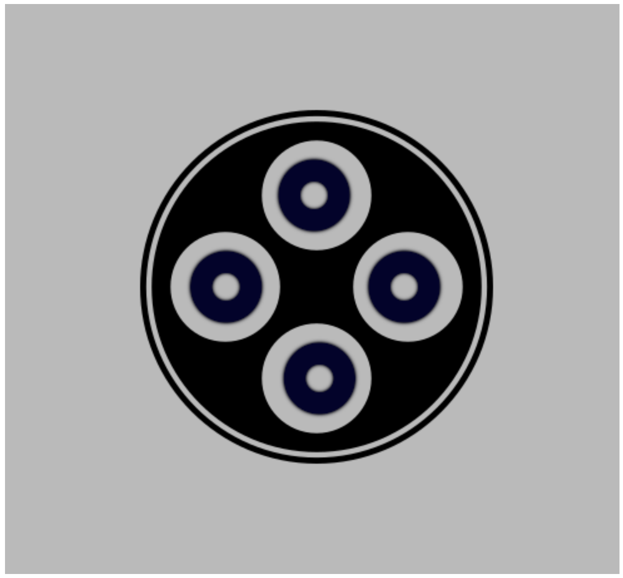


Figure 5. AVT SST stimulus

3.3.5. Fusion Range Test

The fusion range test estimated the ability of individuals to maintain a single fused image in the presence of either horizontal or vertical deviations of the target stimulus in the left/right eye images. Figure 6 shows the horizontal fusion range stimuli (left and right eye images). In order to maintain a fused image, the participant had to cross/uncross their eyes while maintaining a point of accommodation at the display distance. For vertical fusion range, the left/right eye images were displaced in the vertical direction, requiring the eyes to separate in the vertical direction to maintain a single fused image.

Participants were instructed to indicate when the circle became doubled ("broke" fusion) as the circles moved apart. When the direction reversed, participants indicated when the circles returned to a single "fused" image. This task was repeated two times for each of four directions (horizontal crossed, horizontal uncrossed, vertical left eye up, vertical right eye up). This test yielded the following measures: *H Fusion Range* and *V Fusion Range* reported in log arcminutes for the old and new AVT as well as from DRDC.

Figure 6. Horizontal fusion range stimuli (left and right eye images).

3.3.6. OBVA Motion Coherence Test 2.0

Two separate tests were built into the motion coherence test: one that assessed rotational motion perception and one that assessed radial motion perception. During both tests, participants viewed a field of dots presented for 300 milliseconds as shown in Figure 7. A Gaussian envelope characterized the onset and offset of the dot stimuli. During the rotational motion test, dots moved in one of two directions: clockwise or counterclockwise. Participants were instructed to press the right arrow key for clockwise motion and the left arrow key for counterclockwise motion. During the radial motion test, dots moved in one of two directions: expanding (flowing outward toward the edges of the display) or contracting (flowing inward toward the center of the display). Participants were instructed to press the up-arrow key for expansion and the down arrow key for contraction. For both tests, motion coherence (i.e., the percentage of dots moving in a particular direction) was varied according to the Psi adaptive threshold estimation procedure. This test yielded the following measures: *Radial* and *Rotational* and are reported in log proportion coherence. The motion coherence test was developed based on previous research (Morrone, Burr, Di Pietro & Stefanelli, 1999) indicating that there are two cardinal directions for optic flow – radial and rotational. Current military screening tests do not address sensitivity to motion at all, and motion perception may be relevant to predicting operational performance. Although little previous research exists on this topic, some previous research does indicate that motion perception can predict operationally relevant performance (Kruk & Regan, 1983; Kruk, Regan, Beverley, Longridge, & Afhrliot, 1983; Gaska, Winterbottom, & Gray, 2016; Wilkins, Gray, Gaska, & Winterbottom, 2013).

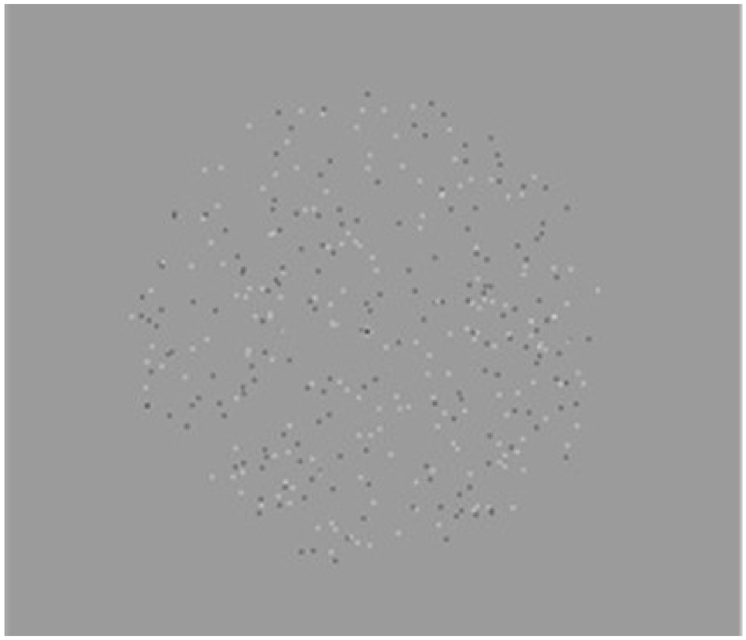


Figure 7. Motion coherence test stimuli.

3.4 Design and Procedures

The study incorporated three principal designs. The first was a within-participants, test-retest study design where the independent variable was test session (one vs. two) and the dependent variable was the test score. The second design was a between-participants design where the independent variable was the particular test battery (old vs. new) and the dependent variable was the test score. The third design was a between-participants design where the independent variable was testing location and the dependent variable was test score. This comparison included only old AVT data as the two locations aside from the OBVA lab only had old AVT software.

3.4.1. Comparing the old and new AVT

The new AVT battery was comprised of all the tests included in the old AVT battery with the addition of the Stereo Search test (a newly developed stereo acuity test). The tests in the old and new AVT batteries were taken at identical distances with the exception of the OBVA Cone Contrast Test (OCCT). The OCCT viewing distance changed from four-meters to one-meter in 2018. The old AVT includes results from both distances, and the new AVT only includes OCCT data at one-meter.

See Table 1 to compare both stimulus display time and time between trials for each old and new AVT test. Stimulus display time is the amount of time the test stimulus appeared on screen before disappearing. If participants did not respond during the stimulus display time, the program chose a response at random and continued on to the next trial. Motion coherence was the exception to this, as participants were allowed an unlimited amount of time to respond after the stimulus disappeared. Time between trials describes the duration between one stimulus and the next.

Table 1. AVT Time Settings.

Test	Old		New	
	Stim Display Time	Time Between Trials	Stim Display Time	Time Between Trials
LCST – Contrast Sensitivity	3000	750	8000	750
LCST – Acuity	3000	750	8000	750
OCCT	3000	750	8000	750
Dual Ring – Near	2000	N/A	8000	500
Dual Ring – Far	2000	N/A	8000	500
Stereo Search	N/A	N/A	8000	500
Fusion Range	N/A	N/A	N/A	N/A
Motion Coherence	300 (Gaussian)	unlimited	300 (Gaussian)	unlimited

Note: All times are in milliseconds (ms)
LCST = Landolt C Contrast Sensitivity Test
OCCT = OBVA Cone Contrast Test

The new AVT battery added an instruction screen, demos, and practice for some of the tests as described below. The new fusion range test included a full run-through of the test as practice. The new dual ring tests included an instruction screen (Figure 8) with a demonstration of the test. The demonstration showed both large crossed and uncrossed disparities to ensure that all participants could detect the disparities. Both the old and new AVT included a short practice test of eight trials for the dual ring tests. The new AVT also included an instruction screen with images of large depth disparities, and a short practice test for the stereo search test.

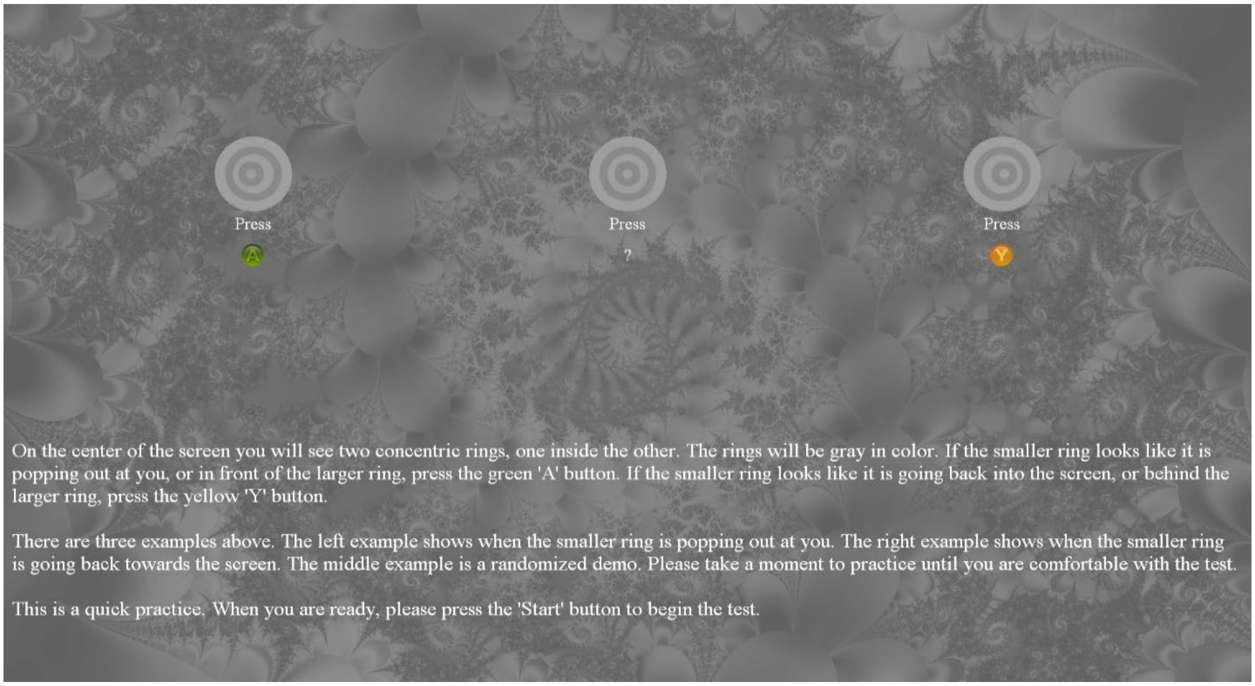


Figure 8. Instruction and demo screen for the near dual ring test in the new AVT battery.

When administering the old AVT in the OBVA lab, tests were repeated in the same experimental session. Session one included LCST and OCCT tests, and session two included the dual ring stereo acuity, fusion range, and motion coherence tests. In comparison, when administering the new AVT in the OBVA lab, all of the tests in the battery were run in the same experimental session and then repeated in the second experimental session. The second session occurred anywhere from one day later to months later, depending on the participant's schedule and lab availability. Therefore, we expect that test-retest reliability could be reduced to some degree for the new AVT administered at the OBVA Lab. Participants completing the old AVT at DRDC completed sessions one and two on separate days, separated by between 1 to 30 days. For the dual ring tests in particular, participants at DRDC ran session one and two in the same week. For the color tests in particular, participants at DRDC ran session one and two within 15 days of one another. DSTG participants completing the near and far dual ring tests completed session one and two back-to-back on the same day.

The new AVT utilized automated data quality checking to ensure that a participant provided acceptable data. Data quality flags were built into each test, which, when activated, would automatically trigger a retest. The retest data overrode their original flagged data so that only the non-flagged data was submitted to the final database. Participants repeated a test up to three times if flags were triggered before their data was submitted to the database. The AVT flagged all threshold and standard error data that was two standard deviations from the mean (based on data from the old AVT). The old AVT lacked these data quality flags. However, as test and retest were conducted in the same session, it was easier to determine how much a participant's scores varied between runs one and two. If a participant's test-retest differences were greater than 0.3 log units, the participant was asked to repeat that particular test.

3.5 Statistical Analysis

Statistical analyses were performed using SPSS version 25 (IBM, Armonk, NY), G*Power 3.1.9.4 (Heinrich Heine Universität Düsseldorf, North Rhine-Westphalia, Germany), and Microsoft Excel (Microsoft, Redmond, WA). Test-retest correlation was analyzed using Pearson's product-moment. A moderate correlation was defined as a Pearson's r value of 0.60 to 0.79 and strong correlation was defined as an r value of 0.80 to 1.00.

Two-tailed independent samples t-tests were used to test for differences between old and new OBVA Laboratory AVT data. A one-way Analysis of Variance (ANOVA) and Tukey's post hoc tests were used to test for differences between DRDC, DSTG, and old OBVA Laboratory near and far dual ring stereoacuity test results. Two-tailed independent samples t-tests were used to test for differences between DRDC and OBVA Laboratory old OCCT L, M, and S cone test results. The statistical significance was set at the alpha = 0.05 level for all statistical tests. When significant differences were identified, effect size was calculated using Cohen's d, where a small effect size was 0.2, a medium effect size was 0.5, and a large effect size was 0.8.

4.0 RESULTS

4.1 Study Sample

At the 711 HPW/RHBC OBVA Lab, a total of 218 and 231 participants completed the old and new AVTs, respectively. Demographic information was not collected for individuals taking the old AVT. Of those participants taking the new AVT, 65 were female and 162 were male (4 unrecorded), ages ranged from 19 to 84 years with a median of 33 years (3 unrecorded), and 103 participants wore vision correction in the form of either glasses or contact lenses (2 unrecorded). One participant was excluded in both the old and new AVTs for providing unreliable and noisy data. Twenty-one participants were excluded from the radial motion test because their thresholds were zero, indicating that they could not tell the difference between expansion and contraction even at 100% coherence. DSTG recruited 42 individuals who completed the far and near dual ring stereo acuity tests. DRDC recruited 129 participants who completed the OCCT and 52 who completed the far and near dual ring stereo acuity test. Of this sample, 37 were female and the subjects' age ranged from 17 to 55 years of age, with a median of 23. Demographic information for the participants in the dual ring stereo acuity test was not recorded.

4.2 Descriptive Statistics

Tables 2 and 3 summarize the descriptive statistics for each of the AVT tests with outliers over two standard deviations from the log mean removed as these outliers are more likely to be poor data than individuals with extremely good or bad vision. The Psi procedure provides an estimate of the threshold error (alpha standard error [SE]), which is a measure of the standard error across trials. That is, on a given trial when the participant is shown the same stimulus level, this measure indicates whether the participant provides the same response; it is a measure of participant consistency. The Psi procedure also provides an estimate of the psychometric function slope at threshold (beta). The higher the psychometric slope, the more sensitive the participant is to the independent variable. See Appendix B for histograms of each test. The OCCT data are collapsed across color normal and color deficient in this table and so while the mean is not a meaningful value, the test-retest and LOA are useful metrics. For descriptive statistics of the color normal data only, see Table 7.

Table 2. Descriptive statistics for each of the AVT tests.

Test	N	Mean Threshold (log Alpha)	Threshold SD	Alpha SE	Slope (Beta)	Test-retest Mean Difference Threshold	Limits of agreement
LCST							
CS 1.25							
Old	151	-0.76	0.21	0.08	0.36	0.02	0.30
New	208	-0.78	0.23	0.08	0.33	0.02	0.43
CS 2.5							
Old	151	-1.28	0.17	0.07	0.4	0.02	0.30
New	212	-1.31	0.19	0.08	0.39	0.01	0.32
CS 16.7							
Old	156	-1.94	0.10	0.06	0.49	0.01	0.18
New	215	-1.96	0.10	0.07	0.47	0.01	0.18
4M Acuity							
Old	144	-0.18	0.08	0.05	0.61	0.01	0.10
New	207	-0.20	0.08	0.05	0.61	0.01	0.12
OCCT							
L cone							
Old	161	-2.18	0.18	0.06	0.47	0.04	0.14
New	167	-2.02	0.45	0.07	0.46	0.02	0.22
DRDC (Old)	129	-1.79	0.53	0.07	0.50	< 0.01	0.19
M cone							
Old	161	-2.14	0.22	0.06	0.44	0.04	0.18
New	165	-2.01	0.30	0.07	0.46	0.03	0.19
DRDC (Old)	128	-1.84	0.41	0.07	0.49	< 0.01	0.21
S cone							
Old	161	-1.12	0.18	0.07	0.34	0.07	0.26
New	166	-1.10	0.21	0.08	0.35	0.04	0.22
DRDC (Old)	127	-1.15	0.17	0.08	0.41	0.01	0.24

DRDC = Defence Research and Development Canada
LCST = Landolt C Contrast Sensitivity Test
OCCT = OBVA Cone Contrast Test

Table 3. Descriptive statistics for each of the AVT tests (continued).

Test	N	Mean Threshold (log Alpha)	Threshold SD	Alpha SE	Slope (Beta)	Test-retest Mean Difference Threshold	Limits of Agreement
Dual Ring Stereo Acuity							
Dual Ring Far							
Old	183	1.87	0.62	0.32	0.27	0.08	0.71
New	218	1.72	0.64	0.34	0.21	0.13	0.98
DRDC (Old)	52	1.93	0.64	0.34	0.23	N/A	N/A
DSTG (Old)	40	2.15	0.59	0.30	0.26	0.07	0.80
Dual Ring Near							
Old	184	1.62	0.63	0.33	0.28	0.06	0.58
New	207	1.40	0.52	0.32	0.27	0.19	0.9
DRDC (Old)	52	1.70	0.72	0.30	0.32	N/A	N/A
DSTG (Old)	42	1.74	0.53	0.32	0.28	0.08	0.78
Stereo Search							
New	214	1.23	0.32	0.15	0.46	-0.11	0.48
Fusion Range							
H Fusion Range							
Old	194	2.20	0.42	N/A	N/A	0.02	0.50
New	217	3.05	0.39	N/A	N/A	0.08	0.64
V Fusion Range							
Old	197	1.37	0.16	N/A	N/A	< 0.01	0.26
New	211	1.35	0.15	N/A	N/A	0.03	0.32
Motion 2.0							
Radial							
Old/New Pooled	251	-0.58	0.25	0.11	0.45	0.02	0.54
Rotational							
Old/New Pooled	266	-0.91	0.18	0.10	0.46	0.01	0.42

DRDC = Defence Research and Development Canada
DSTG = Defence Science and Technology Group

4.3 Test-Retest Reliability, Differences between Versions and Labs

4.3.1. LCST

All session one Contrast Sensitivity (CS) tests were significantly correlated with their session two retests for both the old and new AVT (Table 4, Figure 9). The Pearson product-moment correlation coefficient values showed test-retest correlations were moderate for all tests except the old CS 1.25 and 2.5 tests, which had strong test-retest correlations. The CS 1.25 and 2.5 test-retest correlations were somewhat lower for the new AVT. While correlation characterizes the degree to which two variables are related, Bland-Altman analysis characterizes the differences between two variables (Giavarina, 2015). Limits of agreement (LOA) in this analysis are two standard deviations away from the mean of the difference score, in both directions. Any biases and large differences between variables can also be seen from this analysis. A highly reliable test will result in a mean difference between test 1 and test 2 near zero, and a narrower LOA. Figure 10 shows the Bland-Altman analysis for the LCST.

There were no significant differences between the old and new AVT CS 1.25 tests ($\Delta = 0.02$, $t(357) = 0.82$, $p = 0.42$), CS 2.5 tests ($\Delta = 0.03$, $t(361) = 1.40$, $p = 0.16$), or CS 16.7 tests ($\Delta = 0.02$, $t(369) = 1.732$, $p = 0.08$).

Table 4. Pearson correlation statistics for the contrast sensitivity LCST tests.

		Correlation with Retest		
Test		*N*	*r*	*p*-value
CS 1.25	Old	131	0.81	< 0.01
	New	137	0.63	< 0.01
CS 2.5	Old	130	0.83	< 0.01
	New	138	0.71	< 0.01
CS 16.7	Old	136	0.69	< 0.01
	New	142	0.65	< 0.01

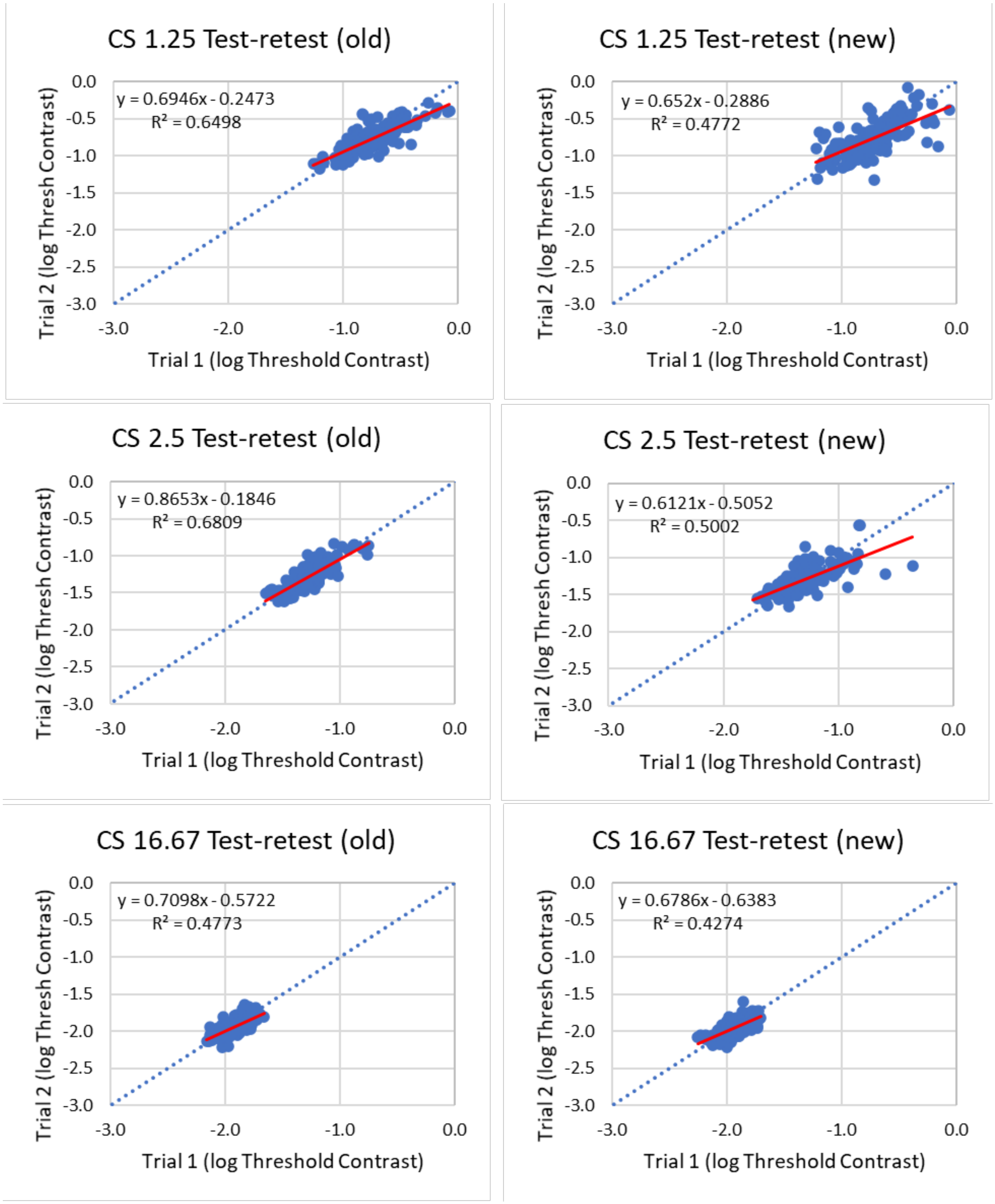


Figure 9. Contrast Sensitivity test-retest scores for OBVA data for contrast sensitivity with a Landolt C stimulus with 1.25, 2.5, and 16.7 arcmin gap sizes. Left: old AVT. Right: new AVT.

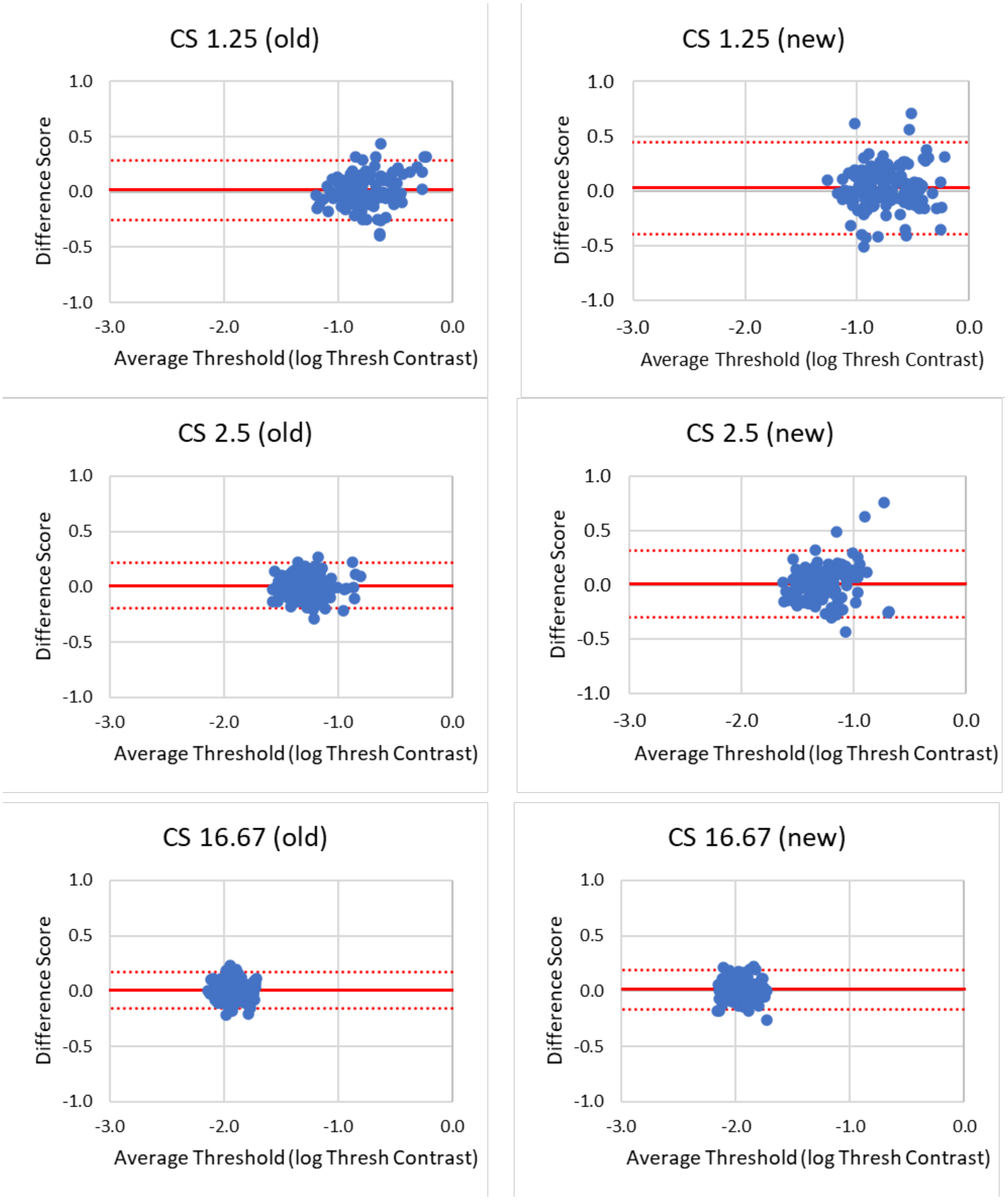


Figure 10. Contrast sensitivity Bland-Altman graphs for contrast sensitivity with a Landolt C stimulus with 1.25, 2.5, and 16.7 arcmin gap sizes. Left: old AVT. Right: new AVT.

Both old and new session one 4M Acuity tests significantly correlated with their session two re-tests (see Table 5, and Figure 11). The Pearson product-moment correlation coefficient values showed test-retest correlations were moderate for both tests. However, the old test showed stronger test-retest correlations than the new test. Figure 12 shows the Bland-Altman analysis for the 4M Acuity tests.

Table 5. Pearson correlation statistics for the 4M Acuity test.

	Correlation with Retest		
Test	*N*	*R*	*p*-value
Old	131	0.78	< 0.01
New	134	0.71	< 0.01

There was a significant difference between the old and new AVT 4M Acuity tests ($\Delta = 0.02$, $t(349) = 2.48$, $p = 0.01$, $d = 0.28$).

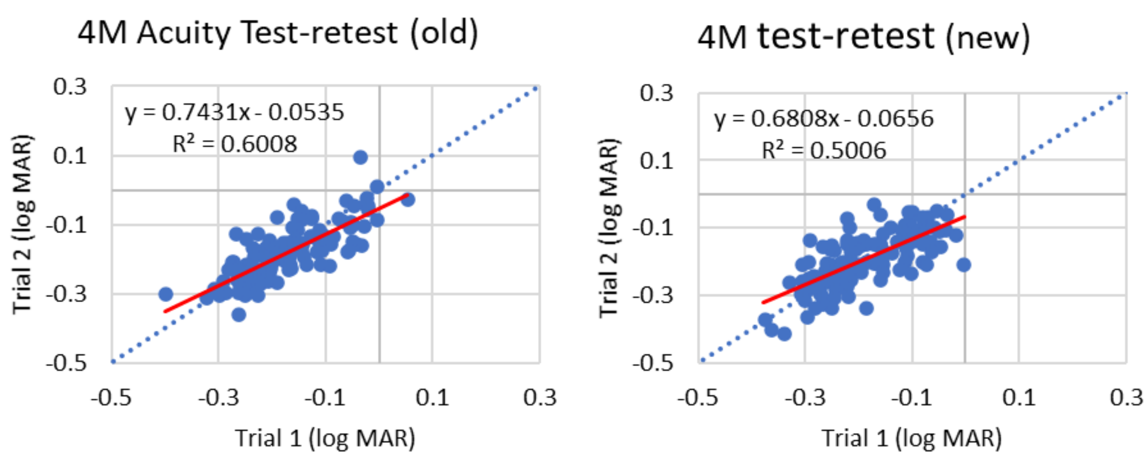


Figure 11. Acuity test-retest scores. Left: old AVT. Right: new AVT.

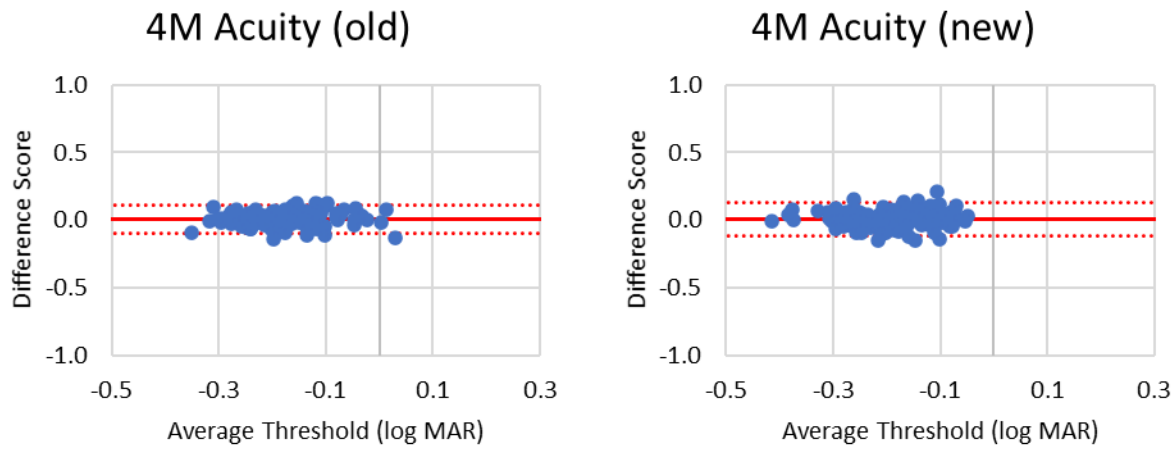


Figure 12. Acuity Bland-Altman graphs. Left: old AVT. Right: new AVT.

4.3.2. OCCT

All session one cone contrast tests (both color normals and color vision deficients included) significantly correlated with their session two retests for the old and new AVT OBVA data and the DRDC data (see Table 6 and Figures 13 to 15). The Pearson product-moment correlation coefficient values showed the test-retest correlations were strong for the L cone and M cone tests and moderate for the S cone test. Figures 16, 17, and 18 show the Bland-Altman analysis for the L, M, and S cone tests, respectively. Again, a narrower LOA in the Bland-Altman analysis means a more reliable test.

Table 6. Pearson correlation statistics for the OCCT tests.

		Correlation with Re-test		
Test		*N*	*R*	*p*-value
L cone	Old	144	0.96	< 0.01
	New	88	0.98	< 0.01
	DRDC (Old)	119	0.98	< 0.01
M cone	Old	145	0.94	< 0.01
	New	87	0.97	< 0.01
	DRDC (Old)	118	0.97	< 0.01
S cone	Old	145	0.75	< 0.01
	New	85	0.81	< 0.01
	DRDC (Old)	117	0.79	< 0.01

DRDC = Defence Research and Development Canada

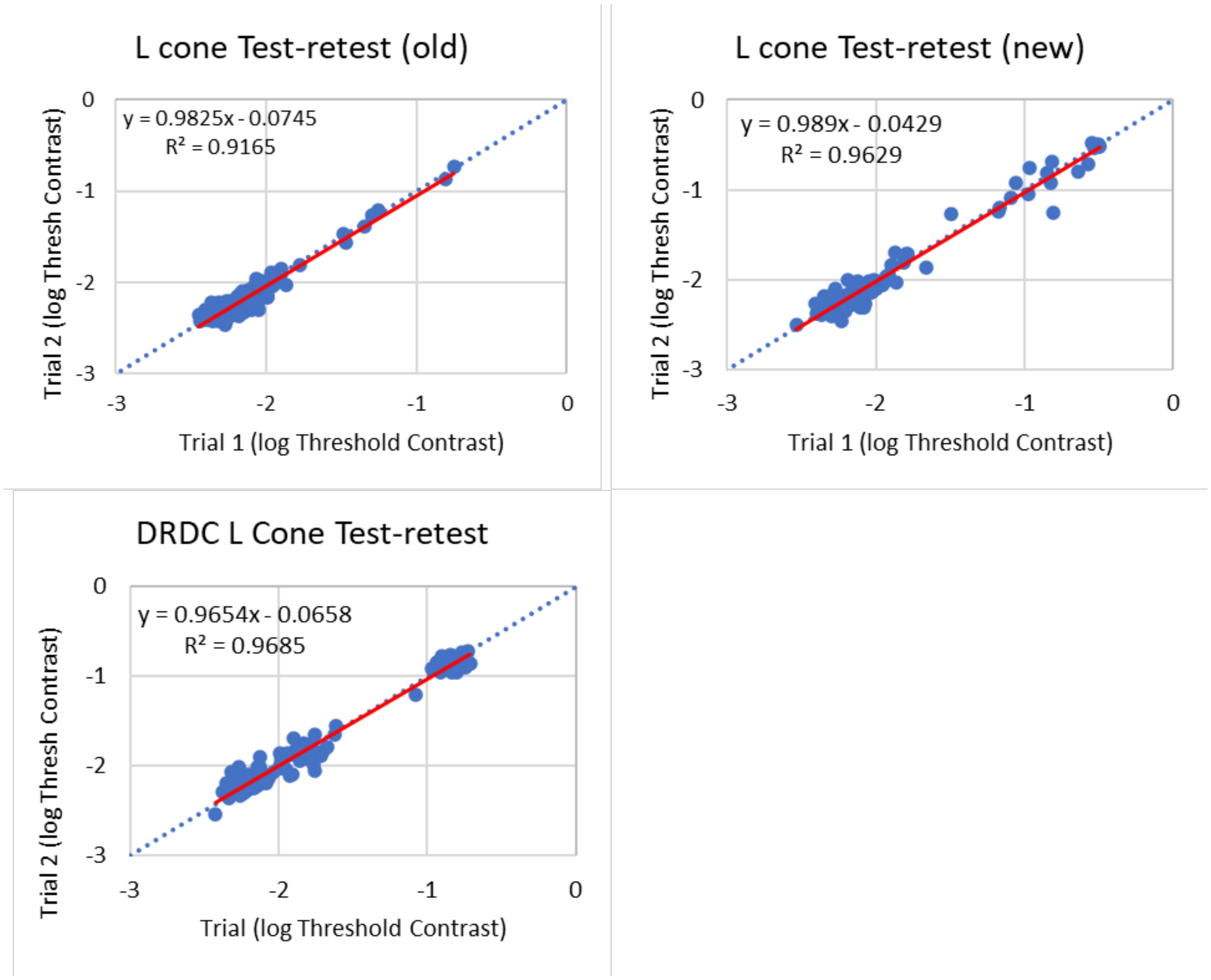


Figure 13. OCCT L cone test-retest scores for OBVA old AVT, OBVA new AVT, and DRDC.

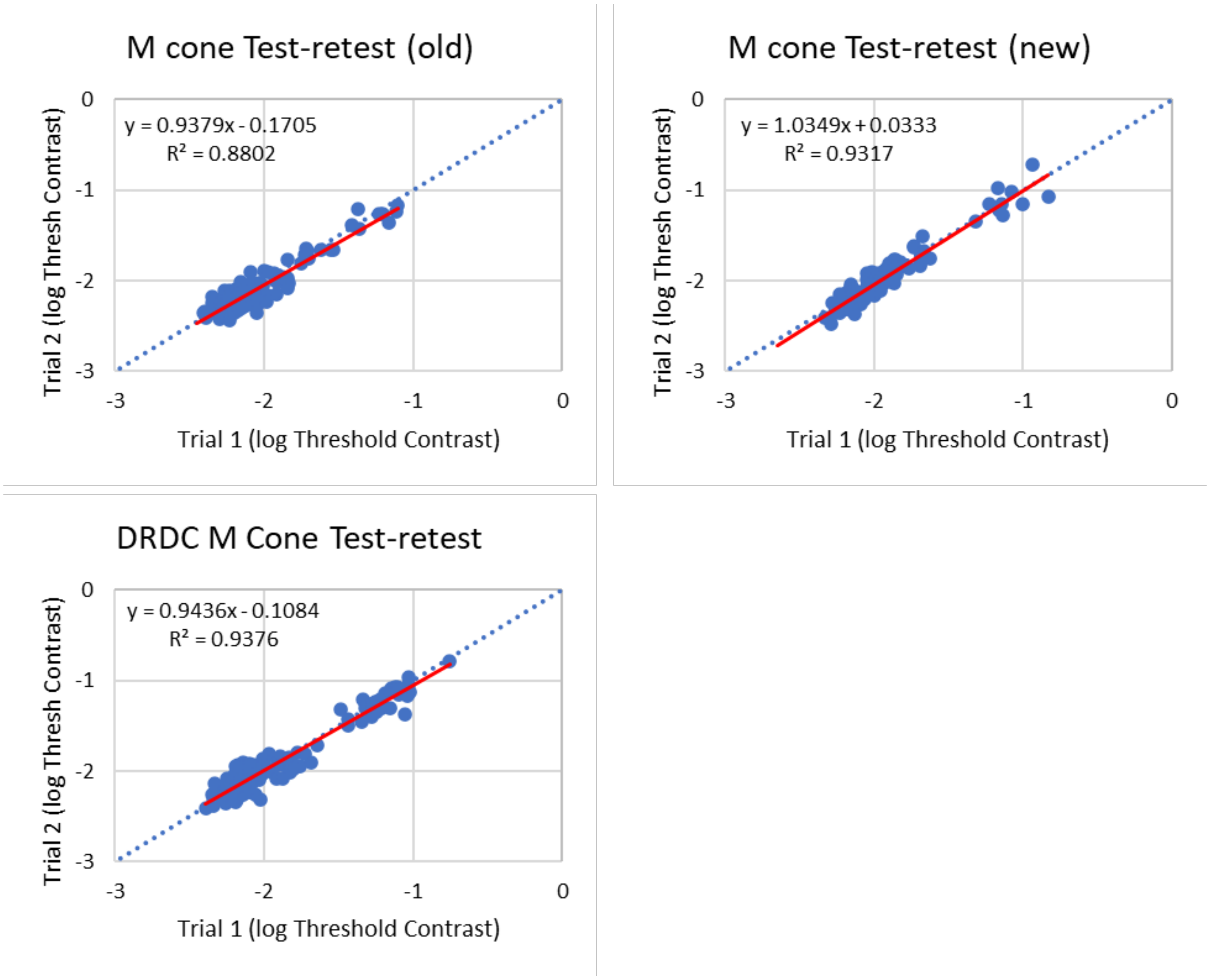


Figure 14. OCCT M cone test-retest scores for OBVA old AVT, OBVA new AVT, and DRDC old AVT.

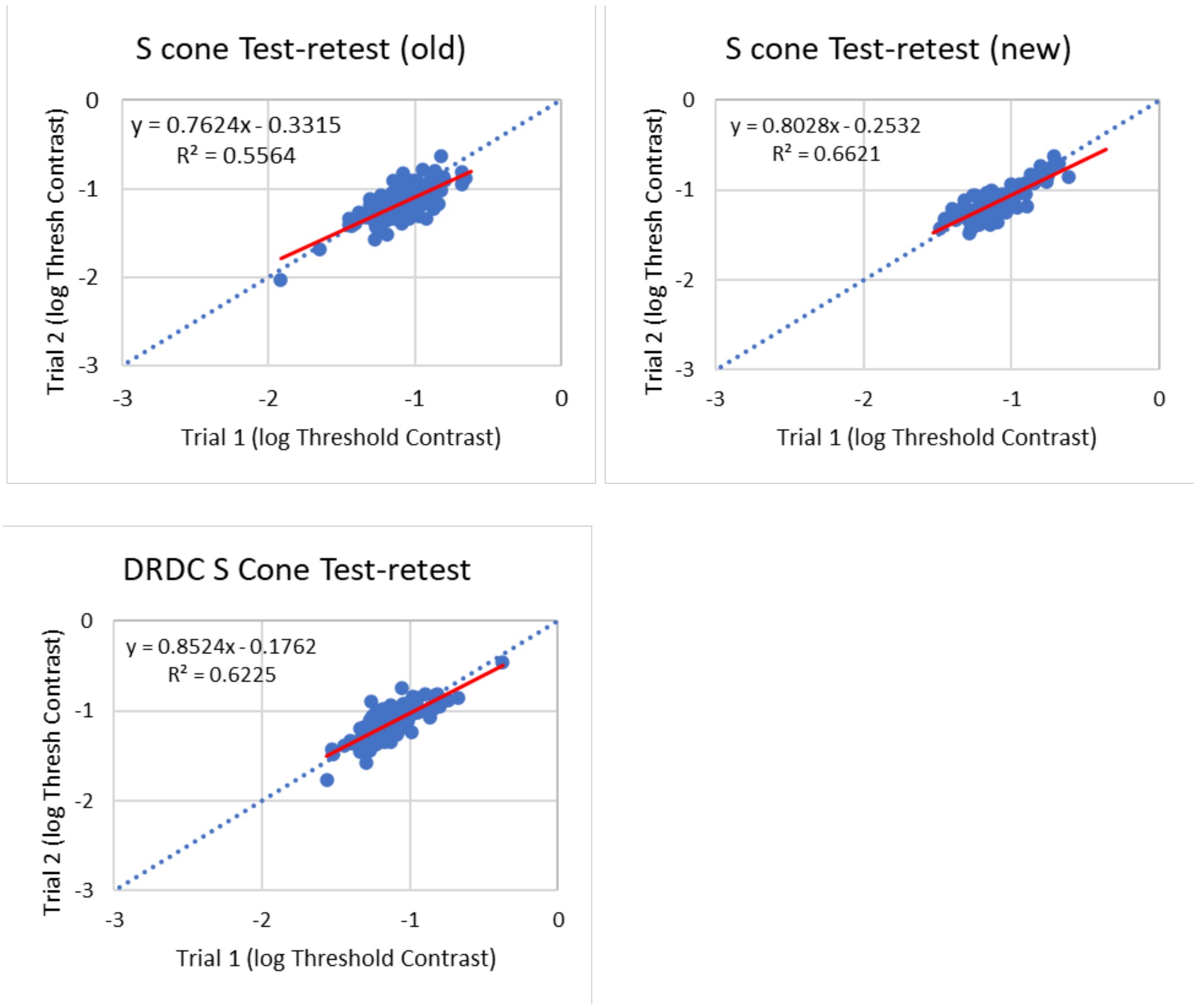


Figure 15. OCCT S cone test-retest scores for OBVA old AVT, OBVA new AVT, and DRDC old AVT.

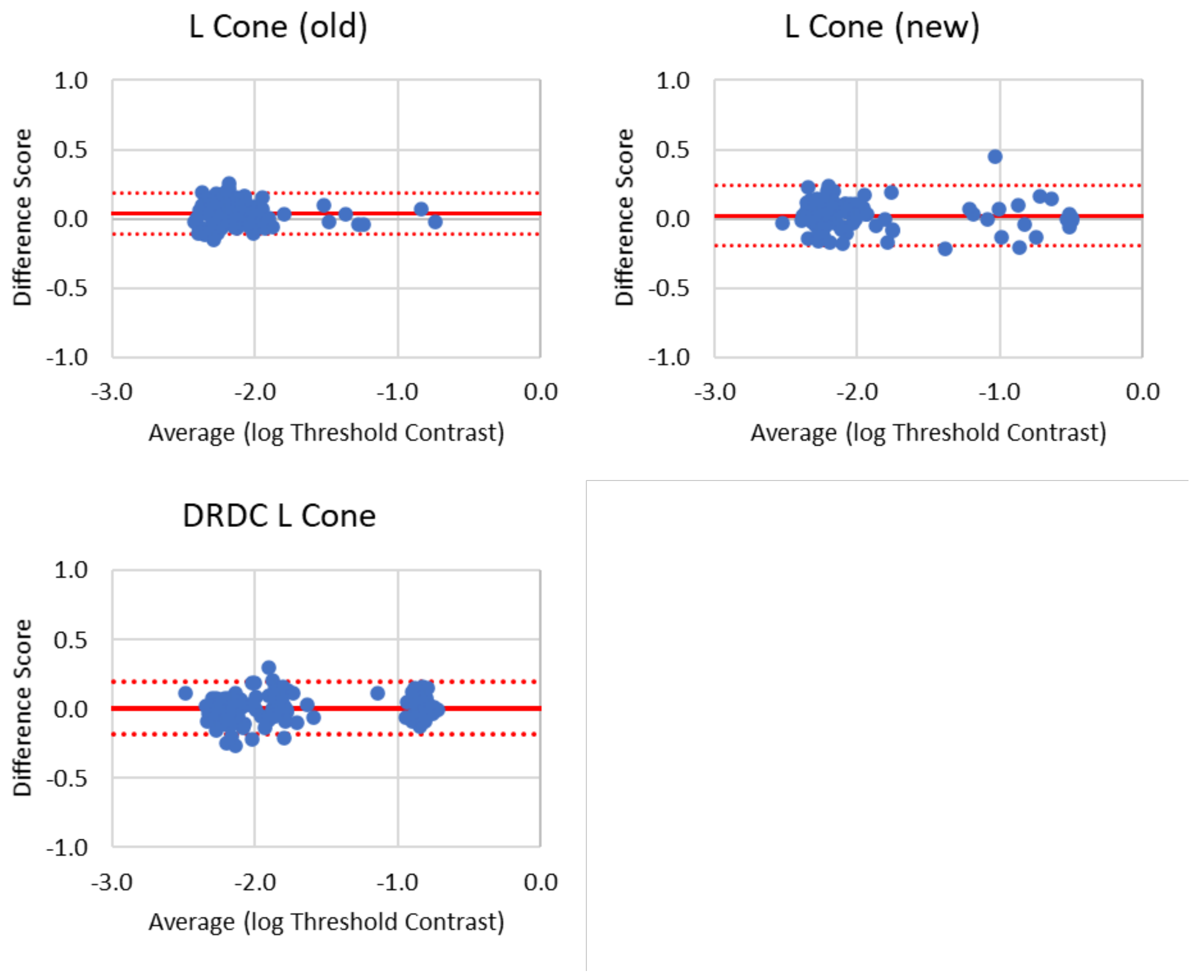


Figure 16. OCCT L cone Bland Altman graphs for OBVA old AVT, OBVA new AVT, and DRDC old AVT.

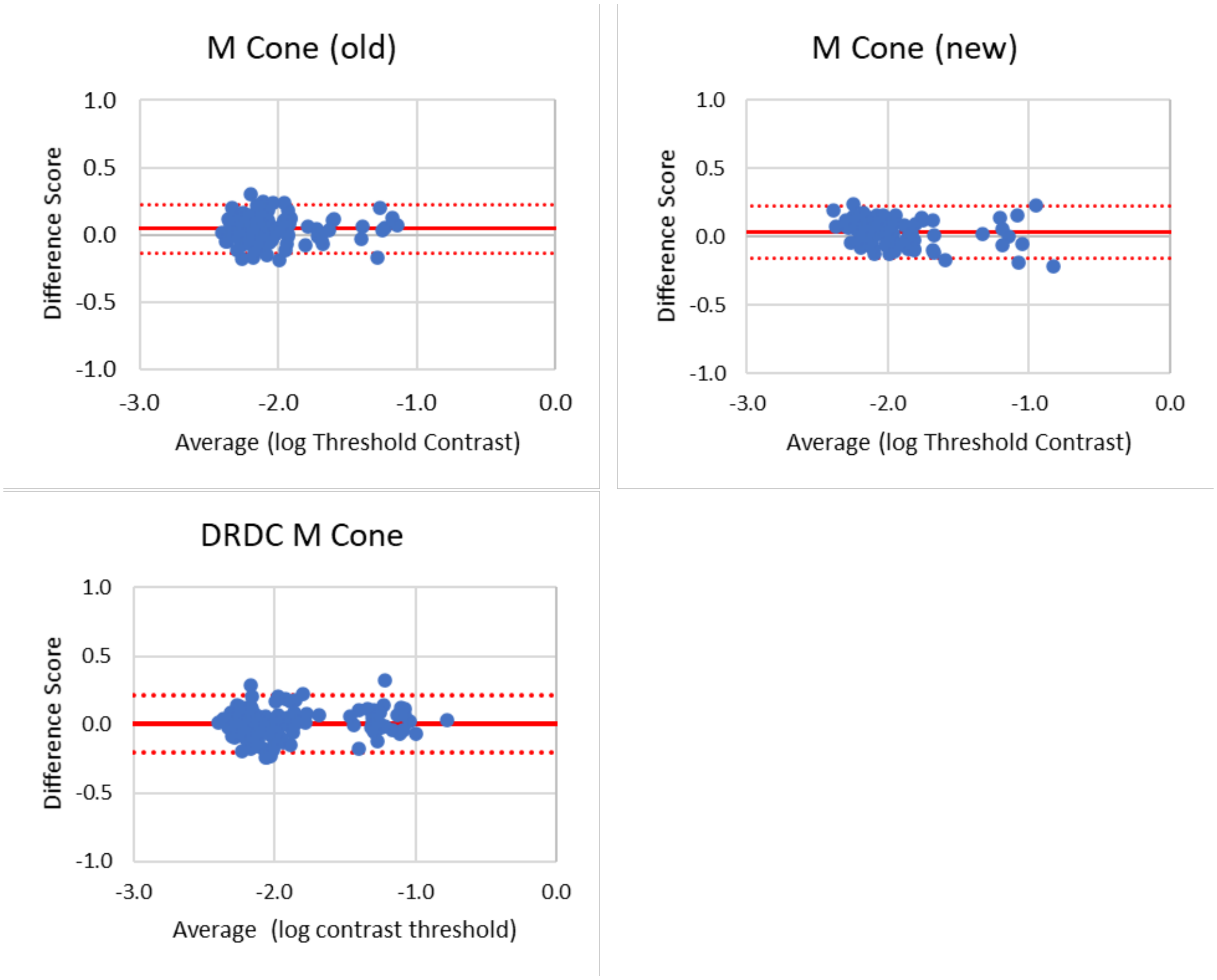


Figure 17. OCCT M cone Bland Altman graphs for OBVA old AVT, OBVA new AVT, and DRDC old AVT.

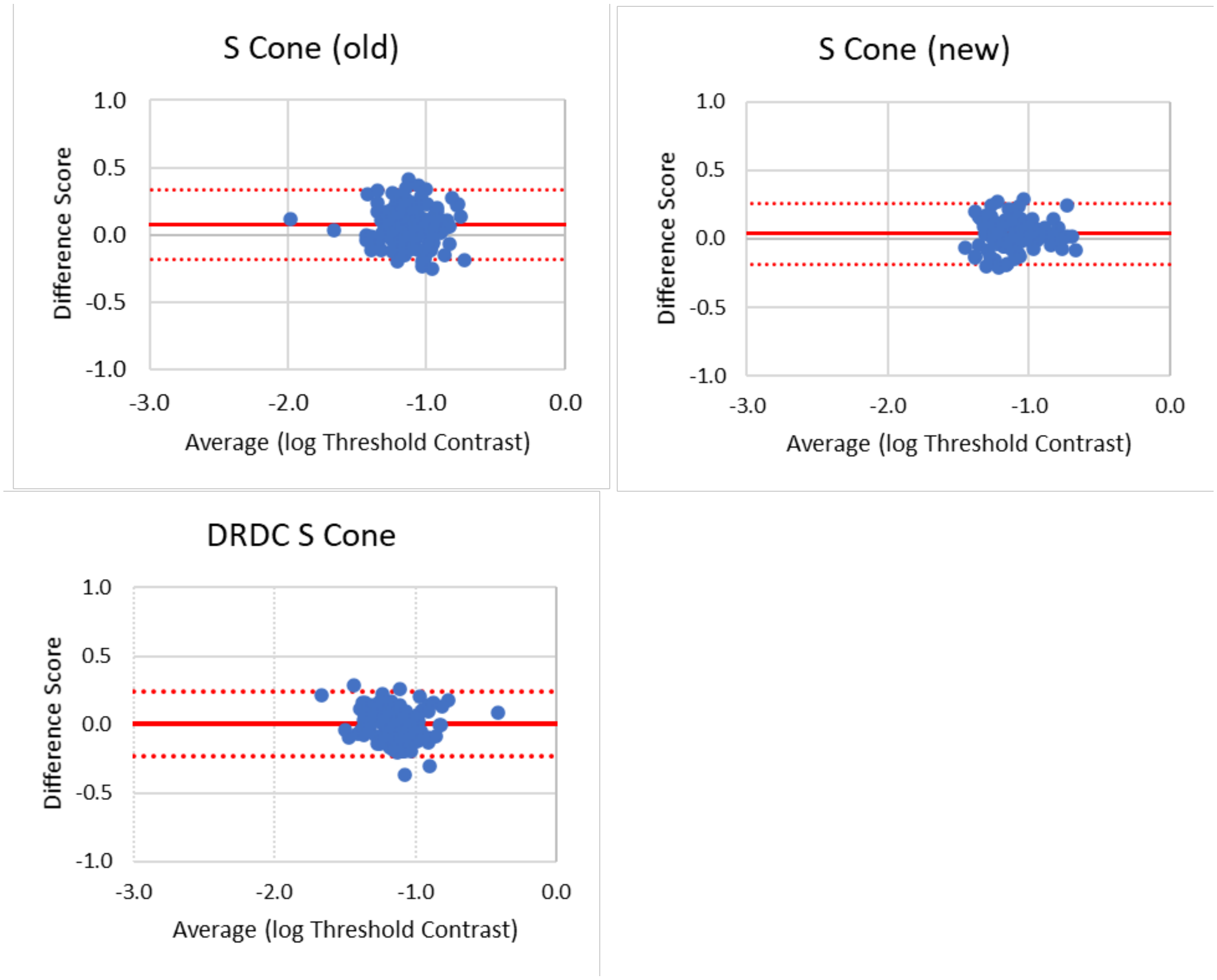


Figure 18. OCCT S cone Bland Altman graphs for OBVA old AVT, OBVA new AVT, and DRDC.

To evaluate differences between test versions and between labs, data from color vision deficient participants were removed because each lab used different recruiting strategies – DRDC recruited approximately equal numbers of color normal and color deficient participants. In general, the OBVA Lab recruited randomly from the local area except when recruiting for specific color test studies and DRDC used an anomaloscope to categorize subjects during recruitment in order to obtain an even number of color normals and deficients. This report instead defines the two groups using the binocular OCCT scores. Deficient participants were defined as those who scored higher than -1.75 log threshold contrast on the L cone and M cone tests or higher than -0.4 log threshold contrast on the S cone test. The OCCT cutoff of -1.75 was derived using the same criterion as the Rabin monocular CCT pass/fail score used by the USAF until 2018 – 3.7 standard deviations from the color normal average score (see Figure 19; Winterbottom, Gaska, Wright & Hadley, 2017). Descriptive statistics for color normals only are shown in Table 7.

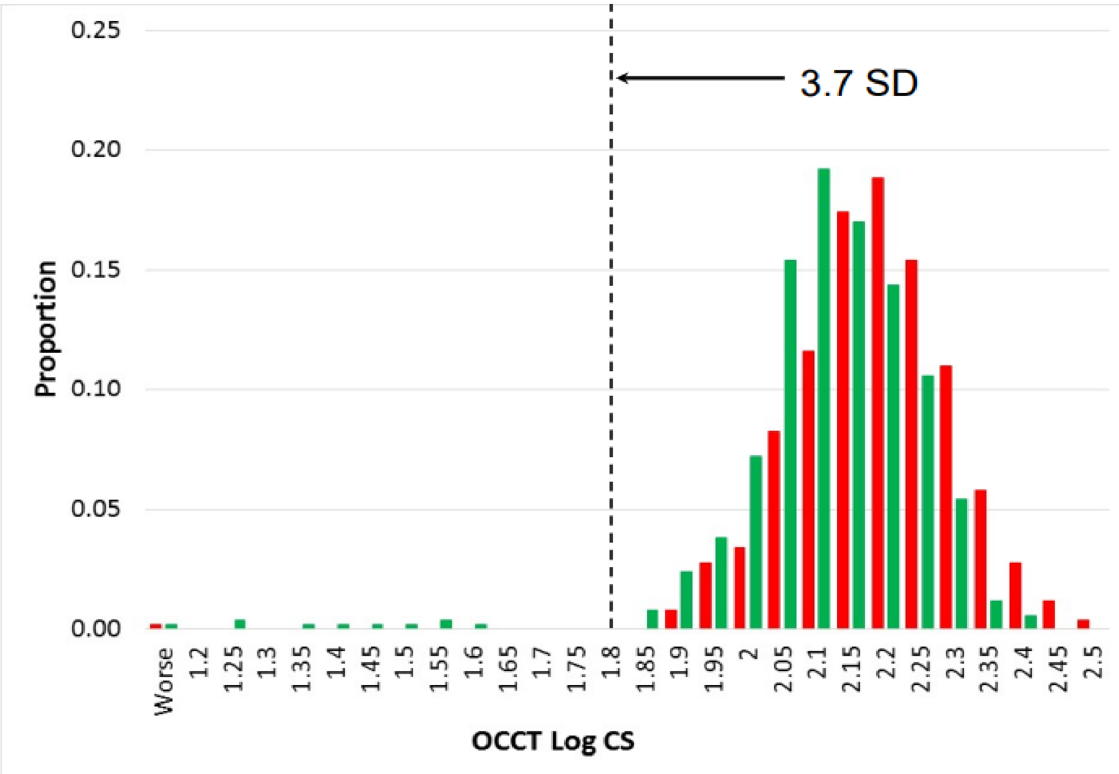


Figure 19. Histogram showing spread of binocular OCCT L cone (red) and M cone (green) scores. Cutoff score for color vision abnormals is 3.7 standard deviations from the mean. From Winterbottom, Gaska, Wright, & Hadley, 2017. International Color Vision Society 24th Symposium.

Table 7. Descriptive statistics for the OCCT tests for color normals only.

Test		N	Mean Threshold (Alpha)	Threshold SD	Alpha SE	Slope (Beta)	Test-retest Mean Difference Threshold	Limits of Agreement
L cone	Old	153	-2.18	0.12	0.06	0.47	0.04	0.14
	New	147	-2.17	0.14	0.07	0.46	0.02	0.20
	DRDC (Old)	97	-2.08	0.17	0.07	0.49	< 0.01	0.20
M cone	Old	145	-2.16	0.13	0.06	0.45	0.04	0.18
	New	144	-2.11	0.14	0.07	0.45	0.05	0.18
	DRDC (Old)	91	-2.08	0.14	0.07	0.48	< 0.01	0.22
S cone	Old	160	-1.12	0.16	0.07	0.34	0.07	0.26
	New	166	-1.11	0.19	0.08	0.35	0.02	0.28
	DRDC (Old)	129	-1.15	0.18	0.08	0.41	0.01	0.28

DRDC = Defence Research and Development Canada

For the OBVA AVT data, there was no significant difference between the old and new L cone tests (Δ = 0.01, $t(298) = -0.51$, $p = 0.61$). The old M cone test had a significantly higher average threshold than the new M cone test (Δ = 0.05, $t(287) = -3.23$, $p < 0.01$, $d = 0.38$). There was no significant difference between the old and new S cone tests (Δ = 0.01, $t(324) = -0.50$, $p = 0.62$).

Comparing between labs, the old L cone test data collected in the OBVA Lab had a significantly lower average threshold than that collected by DRDC (Δ = 0.10, $t(248) = -5.70$, $p < 0.01$, d = 0.71). The old M cone test data collected in the OBVA Lab had a significantly lower average threshold than that collected by DRDC (Δ = 0.08, $t(234) = -4.26$, $p < 0.01$, $d = 0.56$). There was no significant difference between the OBVA Lab and DRDC for the old S cone test (Δ = 0.03, $t(287) = 1.29$, $p = 0.20$).

Though there were significant differences found, they were less than 0.1 log units, which is approximately the same as the threshold standard error for an individual participant. Figure 20 shows the probability density functions for the old and new OBVA, and DRDC LCCT data. Note the large amount of overlap between all of the data.

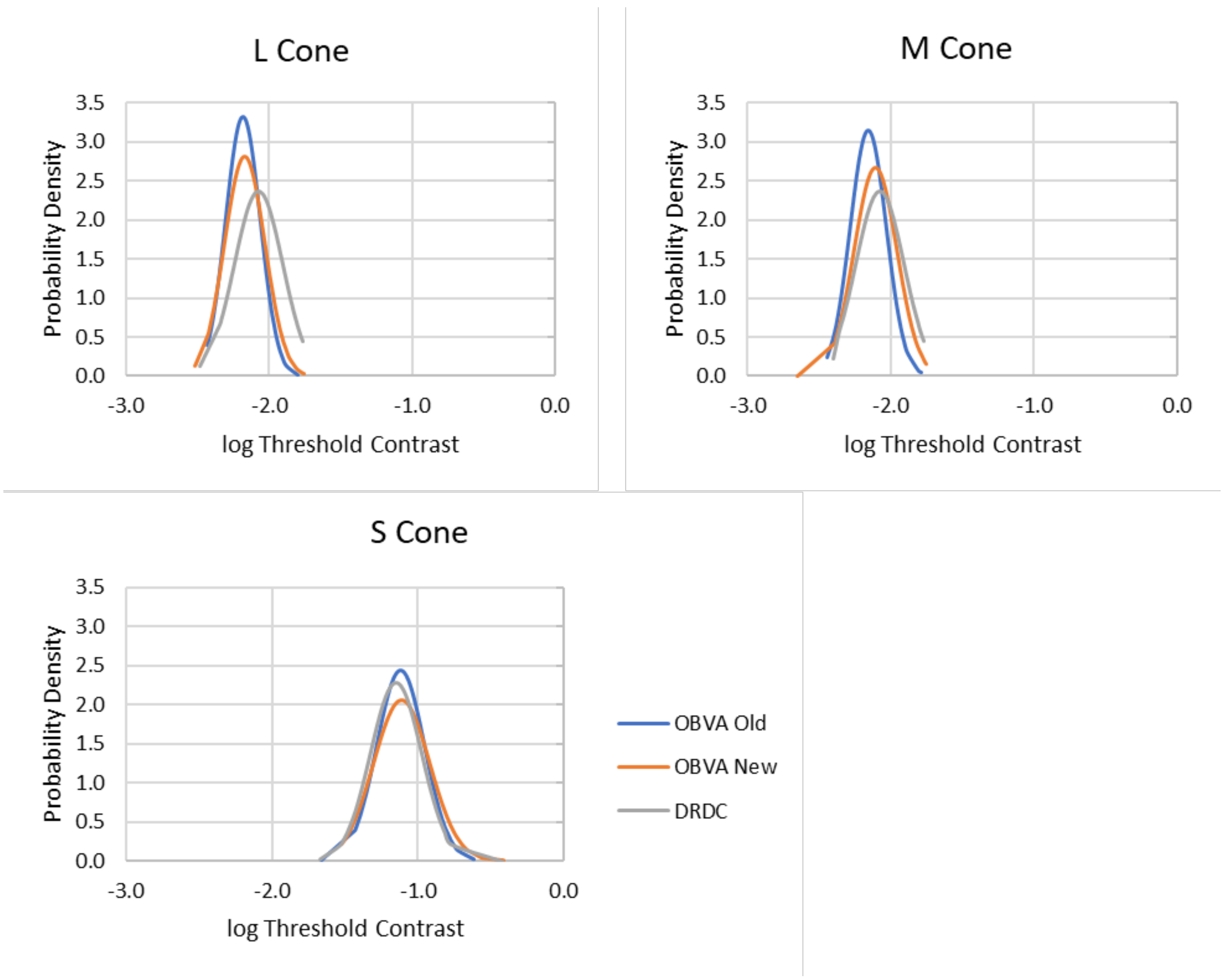


Figure 20. OCCT probability density graphs for OBVA old AVT, OBVA new AVT, and DRDC.

4.3.3. Dual Ring Stereo Acuity

Both session one far and near dual ring stereo acuity tests were significantly correlated with their session two retests for both the old and new AVT OBVA data and the old AVT DSTG data (see Table 8 and Figures 21 and 22). The Pearson product-moment correlation coefficient values showed the test-retest correlations were moderate-to-strong for all tests. However, the old AVT tests showed stronger test-retest correlations than the new tests for both the OBVA Laboratory and DSTG data. The Bland-Altman analysis for the dual ring stereo acuity test is shown in Figures 23 and 24).

Table 8. Pearson correlation statistics for old and new dual ring stereo acuity tests.

		Correlation with Retest		
Test		*N*	*r*	*p*-value
Far	Old	163	0.84	< 0.01
	New	145	0.72	< 0.01
	DSTG (Old)	40	0.77	< 0.01
Near	Old	163	0.90	< 0.01
	New	137	0.68	< 0.01
	DSTG (Old)	41	0.85	< 0.01

DSTG = Defence Science and Technology Group
*Defence Research and Development Canada did not collect test-retest data

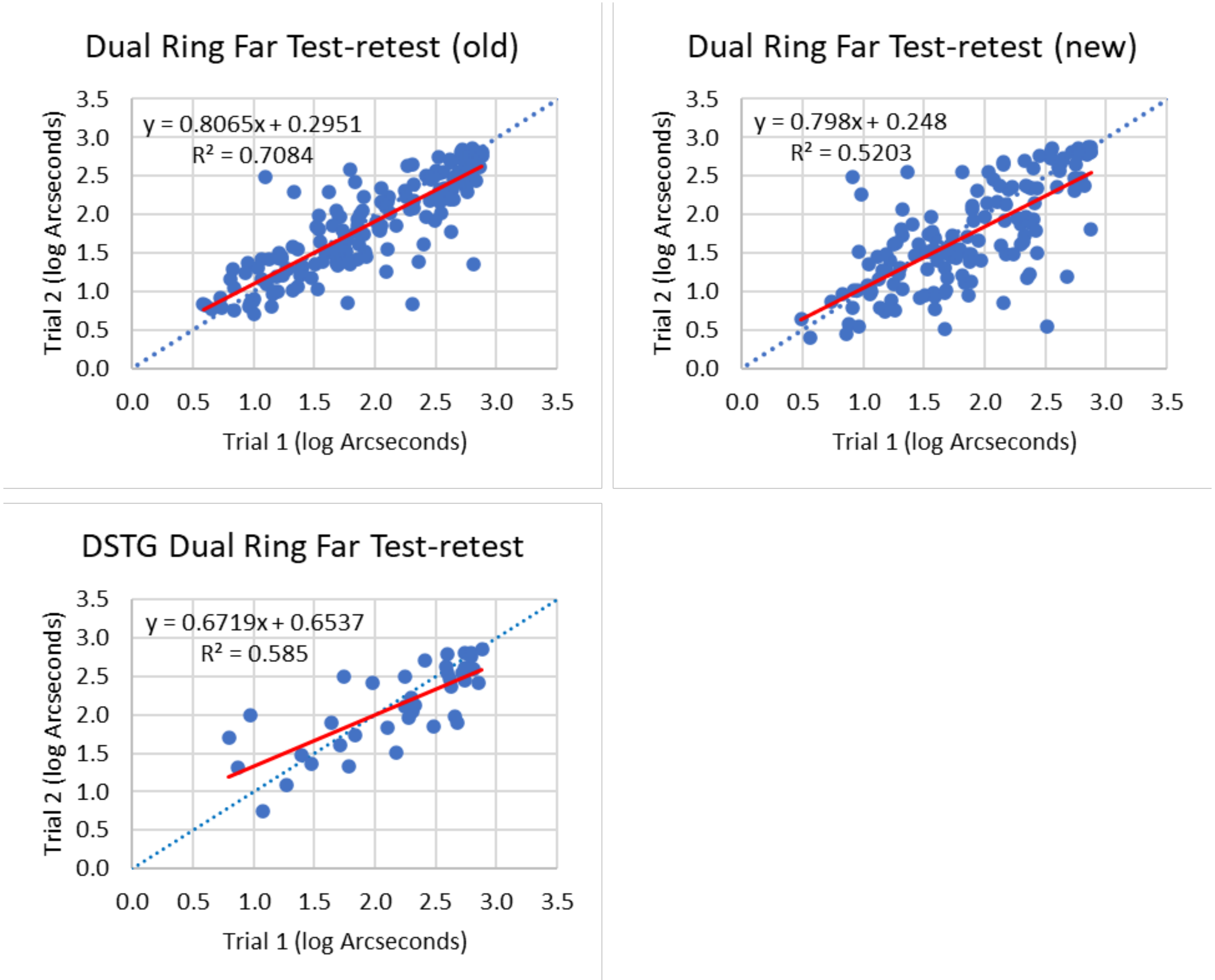


Figure 21. Dual Ring Far test-retest scores for OBVA old AVT, OBVA new AVT, and DSTG old AVT.

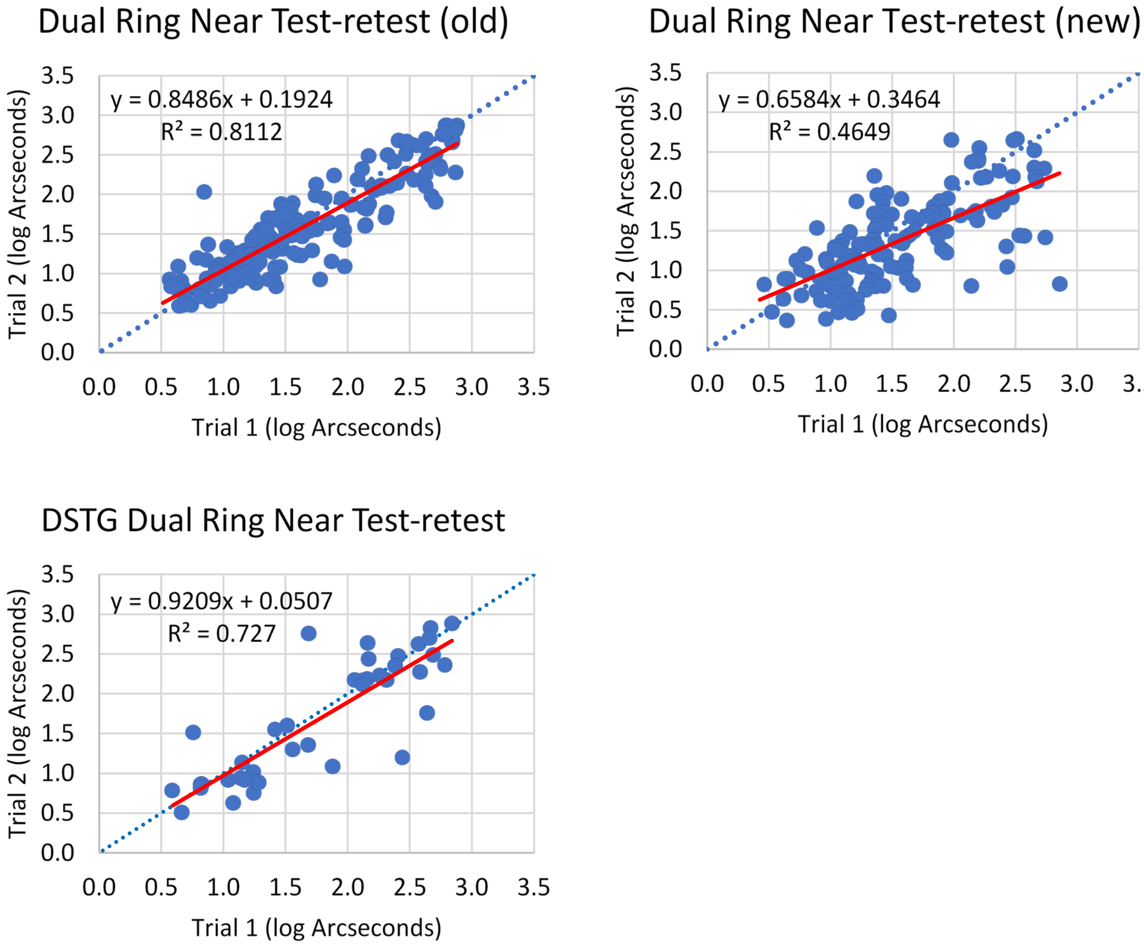


Figure 22. Dual Ring Near test-retest scores for OBVA old AVT, OBVA new AVT, and DSTG old AVT.

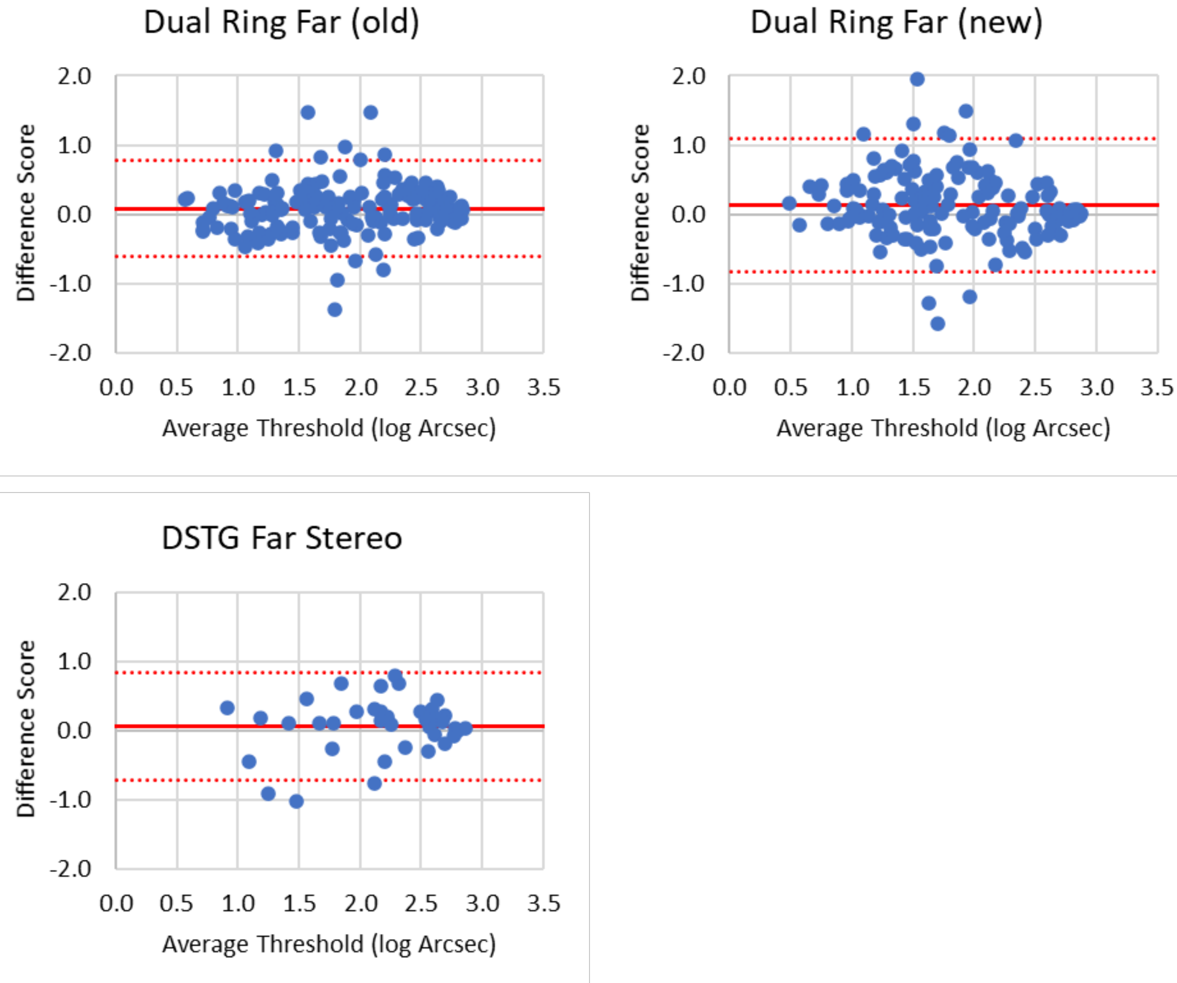


Figure 23. Dual Ring Far Bland Altman graphs for OBVA old AVT, OBVA new AVT, and DSTG old AVT.

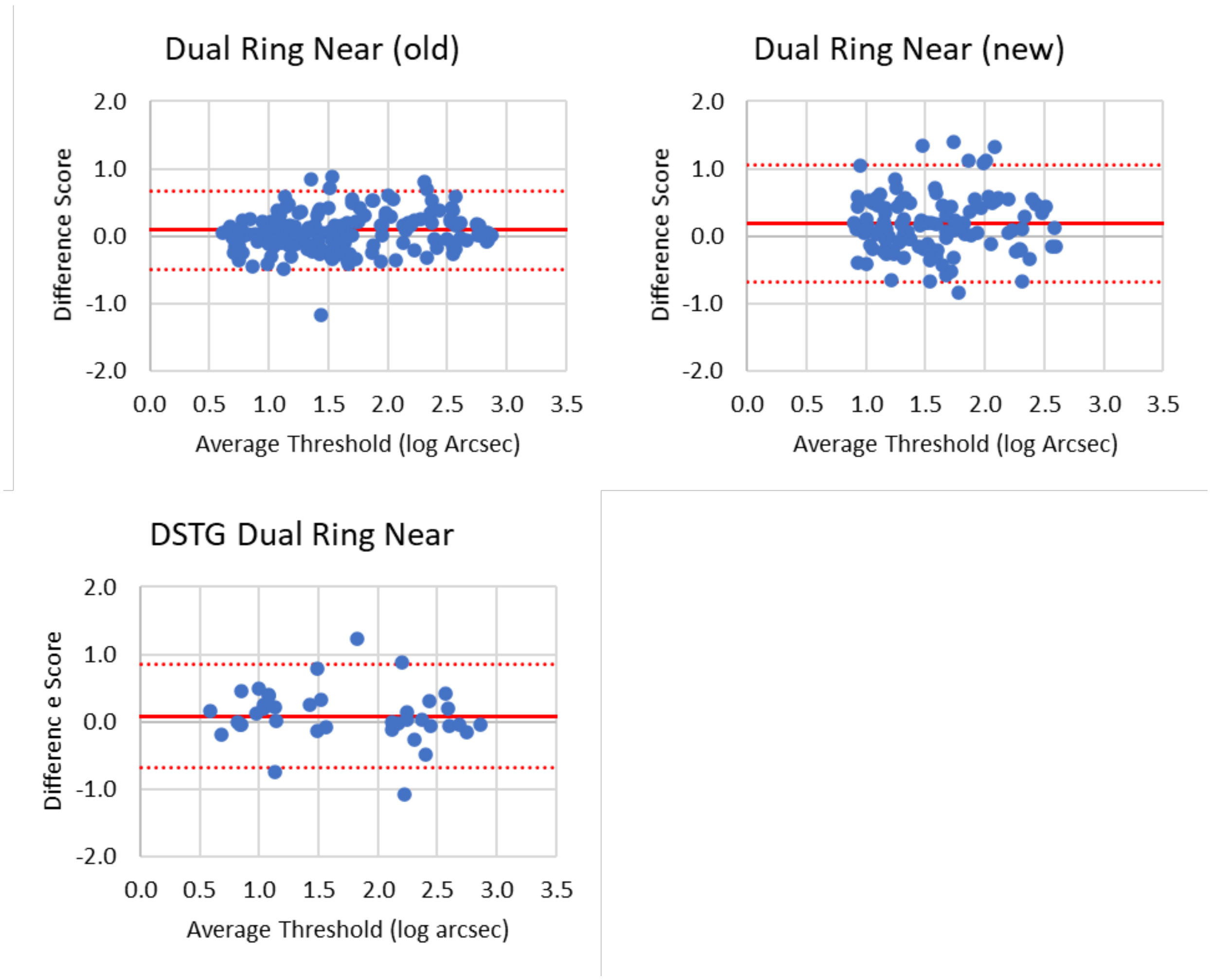
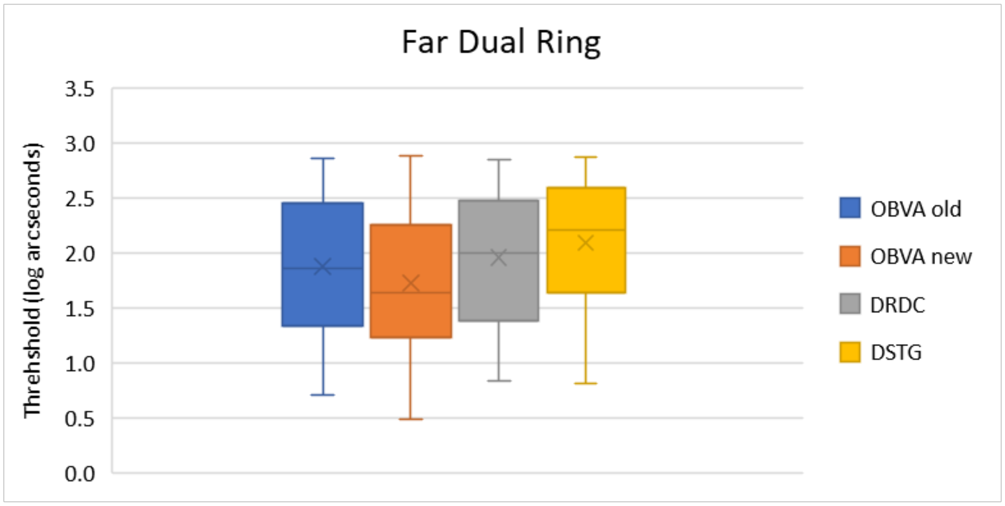


Figure 24. Dual Ring Near Bland Altman graphs for OBVA old AVT, OBVA new AVT, and DSTG old AVT.

For the OBVA Laboratory AVT data, the old far dual ring test average threshold was significantly higher than the new test average threshold ($\Delta = 0.15$, $t(400) = 2.43$, $p = 0.02$, $d = 0.24$). A one-way ANOVA comparing old far dual ring test results across labs showed an effect of location ($F(2, 273) = 3.478$, $p = 0.03$). Post hoc analysis using the Tukey Test indicated the average threshold was significantly higher for the DSTG data versus the OBVA Laboratory data ($\Delta = 0.28$, $p = 0.03$, $d = 0.16$).

For the OBVA Laboratory, the old near dual ring test average threshold was significantly higher than the new test average threshold ($\Delta = 0.21$, $t(391) = 3.61$, $p < 0.01$, $d = 0.36$). A one-way ANOVA comparing old near dual ring test results across labs showed there was no location effect ($F(2, 275) = 0.68$, $p = 0.51$).

Figure 25 shows box and whisker plots for the old and new OBVA, DRDC, and DSTG dual ring data.



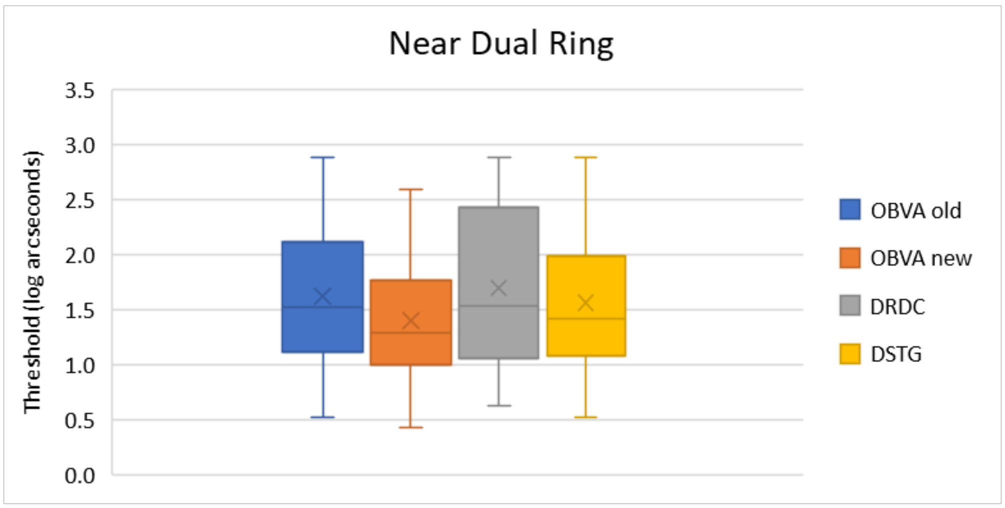


Figure 25. Dual Ring Far and Near boxplots for OBVA old AVT, OBVA new AVT, and DSTG old AVT.

4.3.3.1. Dual Ring Stereo Acuity Test – Effect of Training and Instructions

DSTG researchers deployed the AVT to three different Army bases in Australia and administered the dual ring stereo acuity test at different times (N = 70). Between administering the test at each location, the instructions for the test were gradually refined, and additional training for each subject was added. For example, at the second test location, the researchers introduced a 3D printed mock-up of the appearance of the two rings with the inner ring popped out in front of the reference ring to illustrate the appearance of the test stimuli, and at the third test location the researchers added a 3D example shown on the monitor. Figure 26 shows average near dual ring stereo acuity (log arc seconds) at each location. A Games-Howell test was used to compare differences across locations, which showed that test scores at location 3 were significantly lower (better) than those at locations 1 (mean difference = 0.689, p = 0.002) and 2 (mean difference = 0.477, p = 0.002). Similarly, the new AVT incorporated additional written and auditory instructions, and included an example of the appearance of the dual ring stimulus with a large disparity before the test began (see Figure 8).

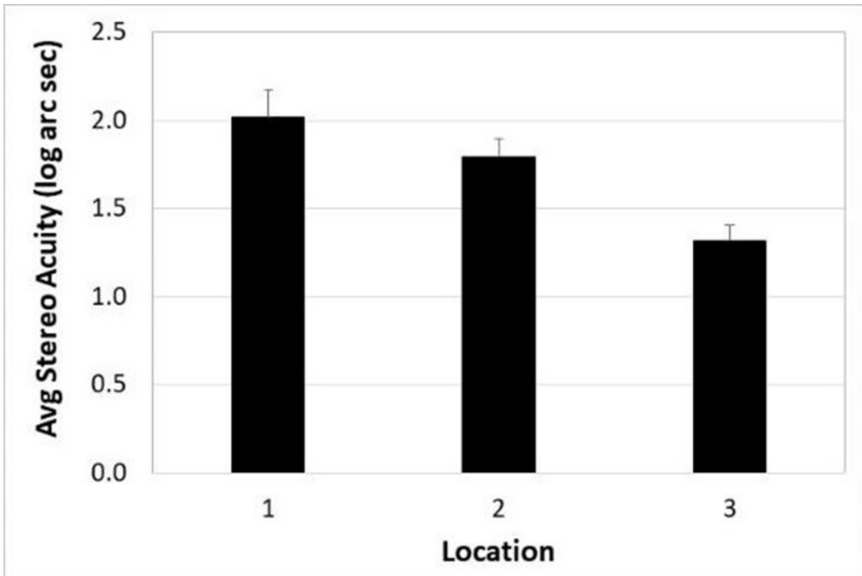


Figure 26. Average near dual ring stereo acuity for subjects at three different Army bases in Australia. Dual ring test instructions and training were refined for each successive location.

4.3.4. Stereo Search

Session one Stereo Search significantly correlated with session two Stereo Search ($N = 139$, $r = 0.66$, $p < 0.01$) showing a moderate test-retest correlation (Figure 27, left). The Bland-Altman analysis for the Stereo Search test is shown in Figure 27, right.

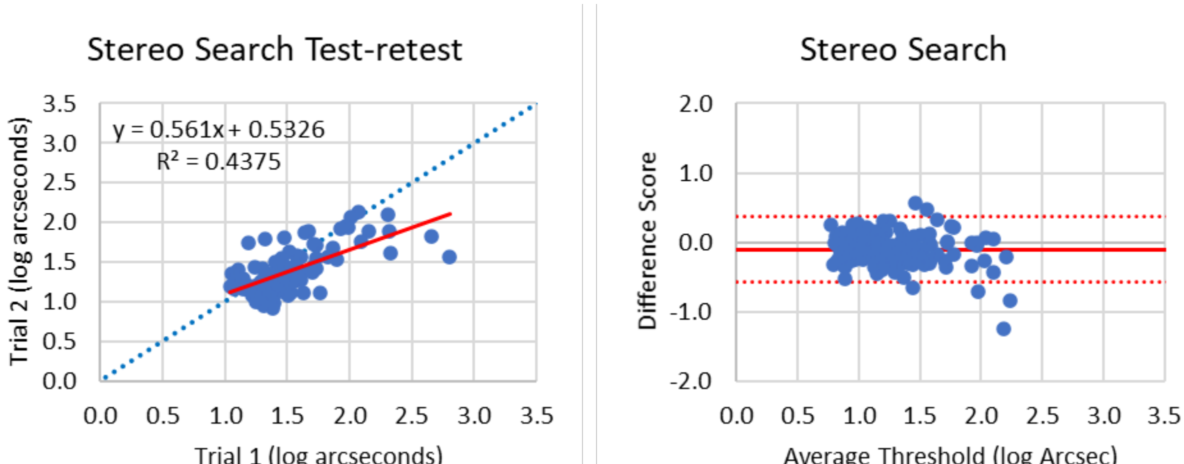


Figure 27. Left: Stereo search test-retest scores. Right: Stereo search Bland-Altman graph

4.3.5. Fusion Range

Both session one horizontal and vertical fusion range tests significantly correlated with their session two retests for the old and new AVT (see Table 9 and Figures 28 and 29). The Pearson product-moment correlation coefficient values show the test-retest correlations were weak-to-strong for all tests. The new and old horizontal OBVA lab results showed similar test-retest correlations, while the new vertical test-retest correlation was higher than the old test. The Bland-Altman analysis for the Fusion Range test is shown in Figures 30 and 31.

Table 9. Pearson correlation statistics for the fusion range tests.

		Correlation with Retest		
Test		*N*	*r*	*p*-value
Horizontal	Old	160	0.84	< 0.01
	New	142	0.82	< 0.01
Vertical	Old	162	0.69	< 0.01
	New	142	0.73	< 0.01

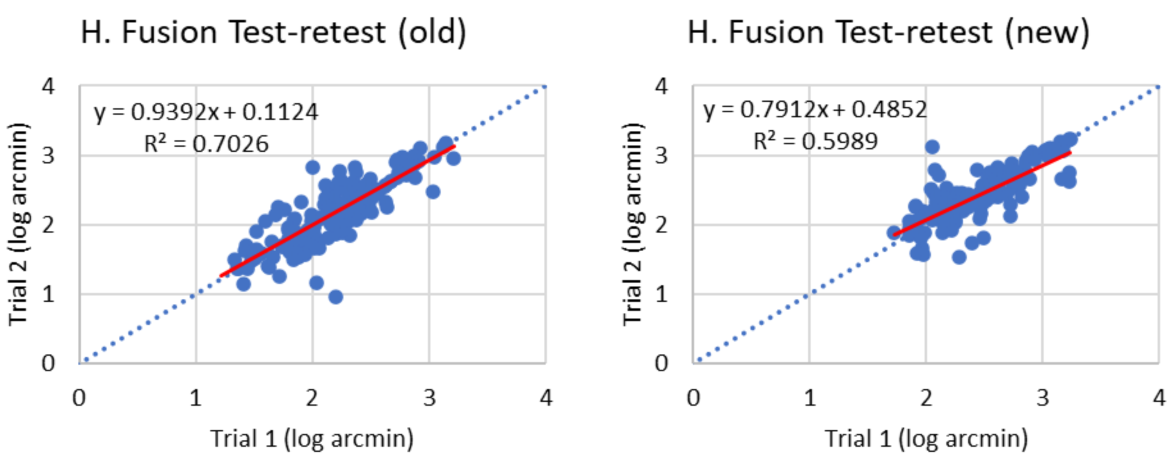


Figure 28. Horizontal fusion range test-retest scores for OBVA old AVT and new AVT

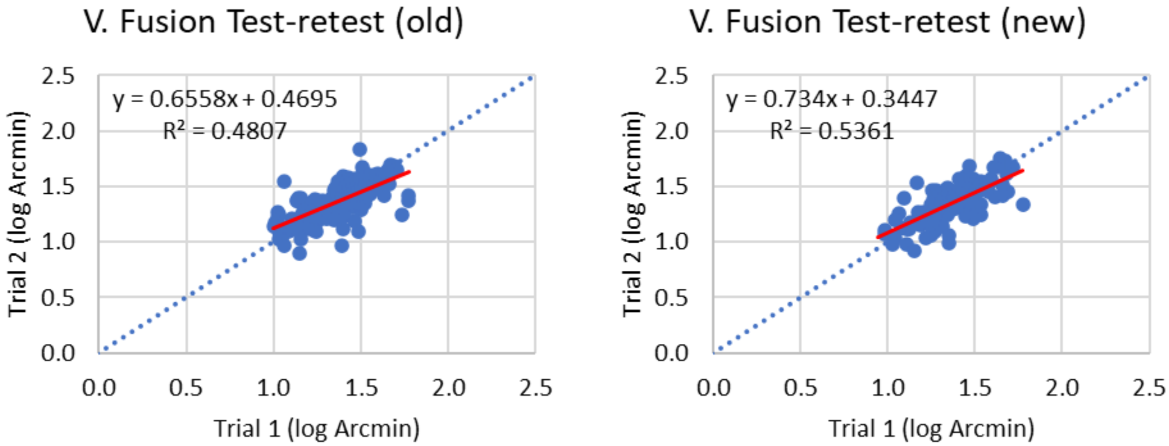


Figure 29. Vertical fusion range test-retest scores for OBVA old AVT and new AVT

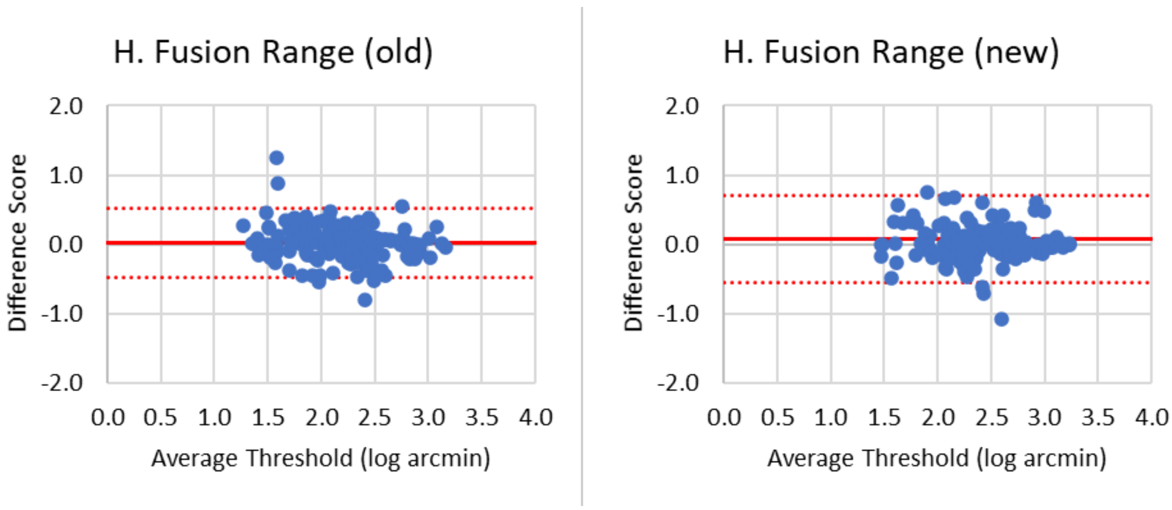


Figure 30. Horizontal fusion range Bland-Altman graphs for OBVA old AVT and new AVT.

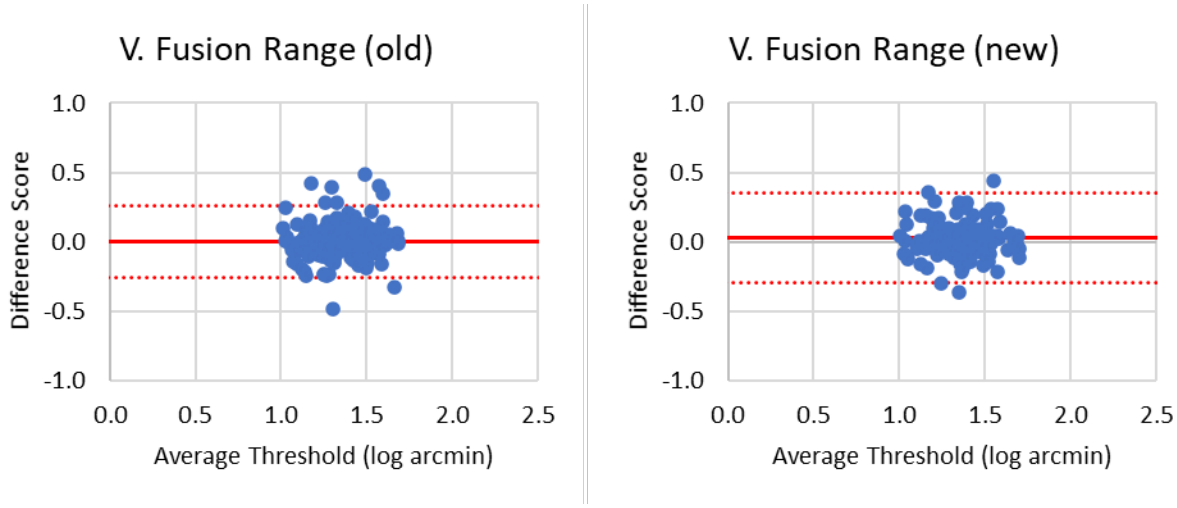


Figure 31. Vertical fusion range Bland-Altman graphs for OBVA old AVT and new AVT

The new average horizontal fusion range was significantly higher than the old range ($\Delta = 0.20$, $t(409) = -5.01$, $p < 0.01$, $d = 0.49$). There was no significant difference between the old and new average vertical fusion ranges ($\Delta = 0.02$, $t(407) = 1.21$, $p = 0.23$).

4.3.6. Motion Coherence

Both session one radial and rotational motion coherence tests significantly correlated with their session two re-tests (see Table 10 and Figure 32). The Pearson product-moment correlation coefficient values showed test-retest correlations were moderate for both tests. The Bland-Altman analysis for the Motion Coherence test is shown in Figure 33. Approximately 12% of the subjects who completed a test and retest for the radial motion test were unable to detect the radial motion even at 100% coherence for one or both of the tests (i.e., obtained a log coherence score of 0). The data for those subjects has been excluded for the purposes of this analysis. Notably, a similar limitation was not observed for the rotational test.

Table 10. Pearson correlation statistics for the motion coherence tests.

	Correlation with Retest		
Test	*N*	*R*	*p*-value
Radial	168	0.50	< 0.01
Rotational	171	0.51	< 0.01

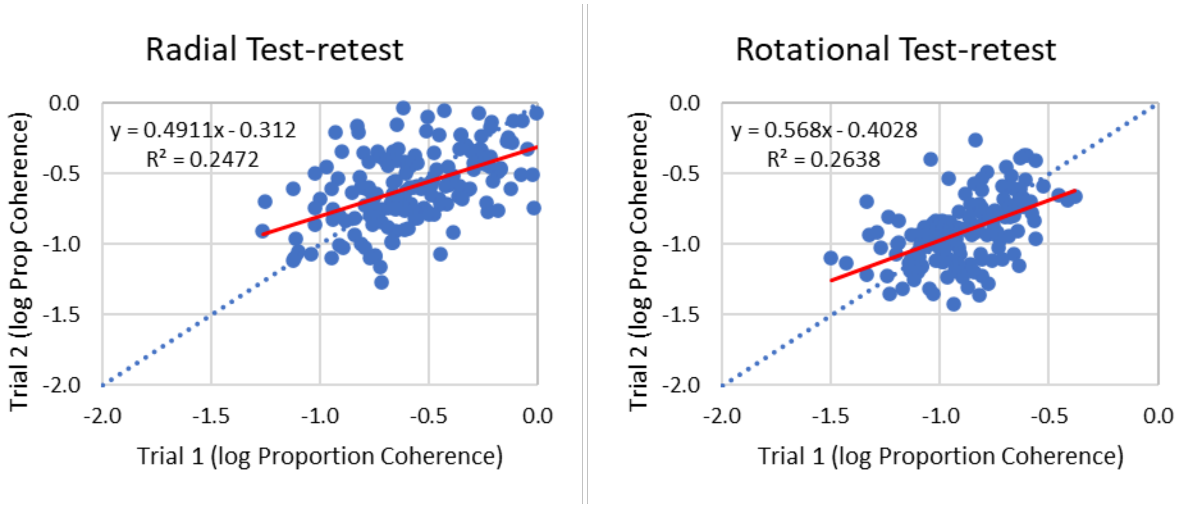


Figure 32. Motion test-retest scores.

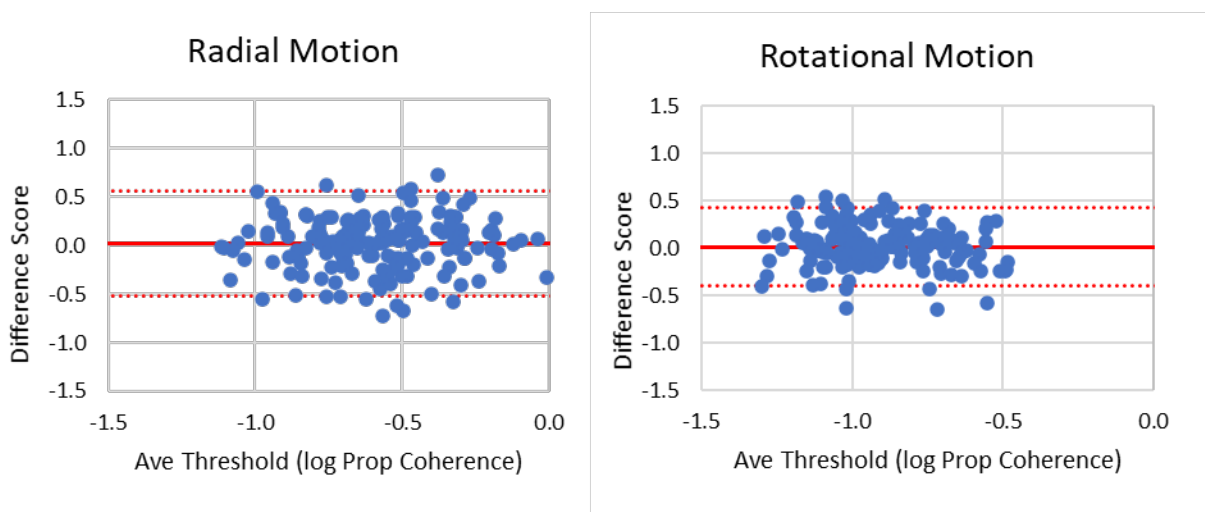


Figure 33. Motion Test Bland-Altman graphs

5.0 DISCUSSION

The objectives of the research summarized in this report were:

1. Evaluate test-retest reliability of each test and feasibility of each test for use as a standard screening test.
2. Compare results for two versions of the AVT – an older version and an updated version.
3. Examine the consistency of the AVT results at different labs and with different populations.
4. Compare results for two alternative stereo acuity tests (dual ring and stereo search test).
5. Develop a normative database for each of the AVT tests to support development of pass/fail criteria for a new set of modernized, computer-based vision tests for military vision screening.

6. Identify candidate tests suitable for replacing selected existing tests and refine test methods and instructions.

The test-retest reliability of the contrast tests was generally very good. The reliability of the OCCT in particular is excellent (L and M cone limits of agreement approximately 0.2 log units). The other AVT tests were more variable in comparison, but most still had good test-retest reliability with correlations all moderate to strong. The dual ring stereo acuity test was of particular interest and resulted in the lowest reliability (LOA ranging from 0.58 to 0.98 log units depending on test version and location). This will be discussed in more detail below.

Comparison of the test results across versions and labs did result in some statistically significant differences. However, some differences in results are to be expected given differences in subject populations, recruiting methods, changes in test-retest procedure, differing/improved instructions and training/familiarization, and etc. For some tests (e.g., dual ring stereo acuity) it is clear that differences in instructions and practice can have a significant effect on the test results. In fact, the results of the tests are remarkably similar given the differences in sample size, subjects, test environment, test administrators, test version, and variations in instructions.

Table 11 summarizes the differences between old and new versions of the AVT at the OBVA Lab, and differences between the OBVA Lab test results and the DRDC and DSTG test results. As shown, the differences between mean test scores for each test are very small – in all cases smaller than the average standard error estimate of the threshold provided by the Psi algorithm, and much smaller than the limits of agreement. Based on these results, we conclude that the AVT data collected across test versions and labs can be combined to support the development of a normative database and to support decisions concerning pass/fail criteria.

Although the dual ring stereo acuity test has been shown to predict simulated remote vision system air refueling performance (Winterbottom, Lloyd, Gaska, Wright & Hadley, 2016). The results of this research suggest that the stereo search test is a better candidate to support future research and for routine vision screening in a potential commercial version of the AVT. The range of SST scores is much more normally distributed (see appendix), and the test-retest reliability is better (LOA of 0.48 vs. 0.58 to 0.98 for the dual ring).

Table 11. Differences between mean test scores.

	Old vs. New	vs. DRDC	vs. DSTG	Alpha SE (new)	Limits of Agreement (new)
LCST					
CS 1.25	0.02	N/A	N/A	0.08	0.43
CS 2.5	0.03	N/A	N/A	0.08	0.32
CS 16.7	0.02	N/A	N/A	0.07	0.18
Acuity	0.02	N/A	N/A	0.05	0.12
OCCT*					
L-cone	0.01	0.10	N/A	0.07	0.20
M-cone	0.05	0.08	N/A	0.07	0.18
S-cone	0.01	0.03	N/A	0.08	0.28
Dual Ring					
Near	0.22	0.08	0.12	0.32	0.90
Far	0.15	0.06	0.28	0.34	0.98
Stereo Search					
Near	N/A	N/A	N/A	0.15	0.48
Fusion					
Horizontal	0.85	N/A	N/A	N/A	0.64
Vertical	0.02	N/A	N/A	N/A	0.32

*For color vision normal only
DRDC = Defence Research and Development Canada
DSTG = Defence Science and Technology Group
LCST = Landolt C Contrast Sensitivity Test
OCCT = OBVA Cone Contrast Test

5.1 Recommendations for Future Work

The LCST and OCCT tests resulted in excellent test-retest reliability, and the OCCT in particular has already been successfully commercialized. However, some additional work to modify the LCST is needed to add crowding to the visual acuity test, requiring additional validation data. Additionally, a commercial AVT will require the ability to test at truly near distances (14 inches) and at a standard far viewing distance of 20 feet/6 meters. Testing at a viewing distance of 16 inches was not possible with the monitor-based laboratory AVT. Thus, additional research and validation will be required if/when a commercial version is successfully developed, and to develop a normative database for each test at that viewing distance.

Additional work is also needed to identify the preferred stereo acuity test and collect normative data. The SST will be included in current and future research, and the OBVA Lab is also pursuing development of a random dot stereogram stereo test (Posselt, Palmer, Imel, Winterbottom & Hadley, 2021; Posselt & Winterbottom, 2020).

Two of the AVT tests are experimental and are not likely to be standard tests anytime in the near future – the fusion range test and motion perception test. Both tests have been shown to predict operational performance, and a test of ocular alignment/ocular motility is needed to fully characterize visual function. However, although the test-retest reliability is good, the current fusion range test is based on self-reported break and recovery. A redesigned test using objective break and recovery responses is needed before the test could be used as a standard test. The OBVA Lab is pursuing the development of a redesigned fusion range test but is also pursuing evaluation of relatively new commercially available eye-tracking tests in addition to or as an alternative to the fusion range test, such the neuroFit One [https://neurofit.tech] or RightEye [https://righteye.com]. The radial motion test also needs further development. Twenty-one participants were dropped from the results as their test scores indicated they could not discern expanding/contracting motion.

6.0 CONCLUSION

Development of more accurate vision tests to replace 1940's era vision tests is long overdue. "Present military visual standards have existed with little real change since WWII. The design of instruments used to measure visual acuity, color vision, and muscle balance in military clinical settings remains unchanged since the original purchases over 40 years ago" (Moffitt & Genco, 1985). Over 30 years have passed since those authors identified the need to replace the already antiquated vision testing methods. Today, the need is even greater to move beyond simply identifying the presence or absence of disease, particularly as calls for truly occupational-/operational-based medical standards and issues concerning human rights, discrimination, and requests for exception to policy are becoming much more prominent.

To successfully develop new medical screening technology and operational based medical standards requires a large volume of human subject data to establish a normative database to support well-documented and defensible pass/fail criteria. This volume of data is time-consuming and expensive to collect, and a significant challenge for any single laboratory to pursue independently. This research collaboration enabled a cost-effective means to pursue this research to accomplish national objectives concerning military medicine.

The AVT software developed by 711 HPW and shared with the TTCP partner nations proved to be an enabling technology that supported research collaboration at 9 different laboratories/clinics in Australia, Canada, and the United States. The data generated by this collaborative effort contributed to the successful completion of a Cooperative Research and Development Agreement (CRADA) to commercialize a substantially improved version of the Cone Contrast Test developed by Konan Medical, which is now in use worldwide for color vision screening and research. The data generated by this collaborative effort also contributed to the successful launch of Small Business Innovative Research (SBIR) efforts in the United States to develop a commercial version of the automated vision testing (AVT) technology that was central to this research collaboration. The AVT described here provides the basis for a new generation of more accurate and reliable, commercially available, computer-based vision tests to support earlier diagnosis of eye disease, more accurate monitoring of treatment and recovery, and occupational vision screening.

Further, data obtained with the AVT can enable research examining individual differences in visual ability and provide data to examine the relationships between different visual functions. A number of studies attempting to identify the key aspects of vision underlying overall visual capability have been published (e.g., Bargary, Bosten, Goodbourn, Lawrance-Owen, Hogg, & Mollon, 2017; Bosten, Goodbourn, Bargary, et al, 2017; Bosten, Goodbourn, Lawrance-Owen, Bargary, Hogg, & Mollon, 2015; Cappe, Clarke, Mohr, & Herzog, 2014; Halpern, Andrews, & Purves, 1999). However, each study generally uses a unique set of tests, many of which are unlikely to be used for routine vision screening applications (e.g., complex forms test, glass patterns, temporal order judgment, bisection discrimination, etc.). The AVT, with tests deployed across U.S. military services and international, academic, and industry partners to generate a normative database validated against operational task performance, may provide an opportunity to develop an accepted set of computer-based vision tests that can be used to modernize vision screening and can also support research into the aspects of vision underlying visual capability.

7.0 REFERENCES

Abel, L., Karas, R., Douglass, A., Gavrilescu, M., & Gibbs, P. (2019). Decrease in task performance associated with changes in fixations in optically misaligned binocular night vision systems. Presented at the Annual Meeting of the Association for Research in Vision and Ophthalmology (ARVO), Vancouver, BC, Canada.

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. D. (2017). Individual differences in human eye movements: An oculomotor signature?. Vision research, 141, 157-169.

Bosten, J. M., Goodbourn, P. T., Bargary, G., Verhallen, R. J., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. D. (2017). An exploratory factor analysis of visual performance in a large population. Vision Research, 141, 303-316.

Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., Hogg, R. E., & Mollon, J. D. (2015). A population study of binocular function. Vision Research, 110, 34-50.

Cappe, C., Clarke, A., Mohr, C., & Herzog, M. H. (2014). Is there a common factor for vision?. Journal of vision, 14(8), 4-4.

Deas, L., Allison, R., Hartle, B., Irving, E., Glaholt, M., & Wilcox, L. (2017). Estimation of Altitude in Stereoscopic-3D Versus 2D Real-world Scenes. Proceedings of the IS&T International Symposium on Electronic Imaging 2017, Stereoscopic Displays and Applications XXVIII.

Douglass, A., Karas, R., Gavrilescu, M., White, D., Gibbs, P., & Abel, L. (2018). Decrease in gaze time to peripheral symbology in optically misaligned binocular night vision systems. Presented at the Annual Meeting of the Association for Research in Vision and Ophthalmology (ARVO), Honolulu, HA, USA.

Gaska, J., Winterbottom, M., Fu, J., Eisenhauer, G., & Hadley, S. (2019). Operational Based Vision Assessment: Color Deficiency and Performance on Simulated Remotely Piloted Aircraft Color-Coded Display Tasks. Presentation at the Aerospace Medical Association Annual Meeting, Las Vegas, NV.

Gaska, J., Winterbottom, M., Bullock, C., & Hadley, S. (2017). The Effect of Visibility and Color Deficiency on PAPI Light Identification in a Simulated Night Landing Scenario. Presentation at the Aerospace Medical Association Annual Meeting, Denver, CO.

Gaska, J., Winterbottom, M., & Van Atta, A. (2016). *Operational Based Vision Assessment Cone Contrast Test : Description and Operation (AFRL-SA-WP-SR-2016-0007)*. Wright-Patterson AFB, OH.

Gaska, J., Winterbottom, M., & Gray, R. (2016). Operational Based Vision Assessment: Visual Motion Sensitivity and Operational performance. Presentation at the Aerospace Medical Association Annual Meeting, Atlantic City, NJ

Gavrilescu, M., Karas, R., Douglass, A., Abel, L., & Gibbs, P. (2019). Changes in physiological parameters induced by optical misalignment in night vision binocular devices. In Situation Awareness in Degraded Environments 2019 (Vol. 11019, p. 110190F). International Society for Optics and Photonics.

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

Halpern, S. D., Andrews, T. J., & Purves, D. (1999). Interindividual variation in human visual performance. Journal of cognitive neuroscience, 11(5), 521-534.

Hartle, B., Sudhama, A., Deas, L. M., Allison, R., Irving, E., Glaholt, M. G., & Wilcox, L. M. (2019). Contributions of stereopsis and aviation experience to simulated rotary wing altitude estimation. Human Factors. DOI: 10.1177/0018720819853479.

Hartle, B., Sudhama, A., Deas, L., Allison, R., Irving, E., Glaholt, M., & Wilcox, L. (2019). Aviation experience and the role of stereopsis in altitude estimation. Presentation at the Aerospace Medical Association Annual Meeting, Las Vegas, NV.

Hovis, J., Almustanyir, A., & Glaholt, M. (2019). Operational Based Color Vision Test Repeatability Indices. Presentation at the Aerospace Medical Association Annual Meeting, Las Vegas, NV.

Hovis, J., Almustanyir, A., & Glaholt, M. (2018). Agreement and predictive values of computer color vision tests with the Holmes-Wright lantern and Farnsworth-Munsell D15. Presentation at the Aerospace Medical Association Annual Meeting, Dallas, TX.

Hovis, J., Almustanyir, A., & Glaholt, M., (2017). A Comparison of Computer-Based Color Vision Tests. Presentation at the Aerospace Medical Association Annual Meeting, Denver, CO.

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

Karas, R., Gavrilescu, M., Douglass, A., White, D., Gibbs, P., Abel, L., Winterbottom, M., & Hadley, S. (2018). Degradation of task performance induced by binocular misalignment. Presentation at the Ramstein Aerospace Medical Summit/NATO Scientific and Technical Organization annual meeting, Ramstein AB, Germany.

Kruk, R., Regan, D., Beverley, K. I., Longridge, T., & Afhrliot. (1983). Flying performance on the advanced simulator for pilot training and laboratory tests of vision. Human factors, 25(4), 457-466.

Kruk, R., & Regan, D. (1983). Visual test results compared with flying performance in telemetry-tracked aircraft. Aviation, space, and environmental medicine.

Morrone, M. C., Burr, D. C., Di Pietro, S., & Stefanelli, M. A. (1999). Cardinal directions for visual optic flow. Current Biology, 9(14), 763-766.

O'Keefe, E., Gaska, J., Bullock, T., Winterbottom, M., & Hadley, S. (2018). Operational Based Vision Assessment (OBVA): Contrast Sensitivity and Air-to-Air Target Detection. Presentation at the Ramstein Aerospace Medical Summit/NATO Scientific and Technical Organization annual meeting, Ramstein AB, Germany.

Posselt, B., Van Atta, A., Winterbottom, M., & Hadley, S. (2021). A new stereo test for military aircrew. Presentation accepted at theAerospace Medicine Association Annual Meeting.

Posselt, B. N., & Winterbottom, M. (2020). Are new vision standards and tests needed for military aircrew using 3D stereo helmet-mounted displays?. BMJ Mil Health.

Prins, N., & Kingdom, F. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes Toolbox. *Frontiers in psychology*, *9*.

Sudhama, A., Hartle, B., Allison, R., Wilcox, L., & Irving, E. (2019). Estimates of simulated ground relief as an operational test of stereoacuity for aviators. Defence Research and Development Canada Contract Report DRDC-RDDC-2019-C119.

Wilkins, L., Gray, R., Gaska, J., & Winterbottom, M. (2013). Motion perception and driving: predicting performance through testing and shortening braking reaction times through training. Investigative ophthalmology & visual science, 54(13), 8364-8374.

Winterbottom, M., Gaska, J., Wright, S., & Hadley, S. (2017, August). *Operational Based Vision Assessment: Cone Contrast Test Research and Development* [Conference presentation]. International Color Vision Society 24th Symposium.

Winterbottom, M., Lloyd, C., Gaska, J., Williams, L., Wright, S., & Hadley, S. (2017). Operational Based Vision Assessment Research: Aircrew Depth Perception Standards and Screening Methods. York University Centre for Vision Research Vision in the Real World International Conference, Toronto, ON, Canada.

Winterbottom, M., Lloyd, C., Gaska, J., Wright, S. & Hadley, S. (2016). Stereoscopic Remote Vision System Aerial Refueling Visual Performance. Proceedings of the Stereoscopic Displays and Applications Conference, San Francisco, CA.

LIST OF ACRONYMS

AFVT	Armed Forces Vision Tester
ANOVA	analysis of variance
AVT	automated vision test
CCT	cone contrast test
CS	contrast sensitivity
DSTG	Defence Science and Technology Group
DRDC	Defence Research and Development Canada
L	long wavelength
LCST	Landolt C contrast sensitivity test
M	middle wavelength
MAR	minimum angle of resolution
OCCT	OBVA cone contrast test
OBVA	Operational Based Vision Assessment
RAAF	Royal Australian Air Force
RCAF	Royal Canadian Air Force
S	short wavelength
USAF	United States Air Force

Appendix A – Instructions for each test included in the AVT Battery

A.1 OCCT

The old AVT instructions were as follows: "Today we will be testing your color vision, contrast sensitivity, and visual acuity. Even though these are three different tests, the instructions are exactly the same. In the center of the screen you will see individual letter C's presented to you one at a time. The letter C's will take one of four orientations: a normal letter C, rotated downward, rotated left, or rotated up. [Note: I usually use my hand to make the shape of whichever orientation of the C that I'm saying so that they understand what the C's will look like.] Your job is to use the controller to identify which direction you see that gap, or that break, in the letter C. If it is a normal letter C the gap will be to the right, if the C is rotated downward, the gap is down, if the C is rotated to the left, the gap is to the left, and if the C is rotated upwards then the gap is up. [Hand them the controller.] We will be using the colored buttons as directional indicators, so the green 'A' button is down, the red 'B' button is right, the yellow 'Y' button is up, and the blue 'X' button is left."

"The first test will be measuring your color vision, so the letter C's will take one of three colors and then various intensities or brightnesses of those colors. Again, just respond to where you see that gap or break in the C. The second test will be measuring your contrast sensitivity, so the letter C's will be various shades of grey and also will randomly vary between three different sizes. The visual acuity test will be similar, but the letter C's will just be plain black. Keep in mind, all of the tests in the lab are adaptive, which means that if you get an answer correct, the next one will be harder, and if you miss an answer, the next one will get easier. The computer will adjust the stimuli so that you get about 80% correct and 20% wrong. This means that it is impossible to get all of them right, so don't get discouraged if you miss a few, it is in no way indicative of you or your performance. In between each of the three tests it will bring you back to this screen where you can rest your eyes if you need to, and then press the 'A' button to cycle through the prompts to begin the next section. Whenever you are ready to begin, press the green 'A' button to cycle through the prompts. The test will start after you see the screen with the big text."

The new AVT instructions were as follows: "This test will be examining your color vision. In the center of the screen you will see the letter C. The C will take one of four orientations: normal letter C, rotated downward, rotated left, or rotated up. Using the round buttons on the game controller, press the button that corresponds with the direction of the gap in the C (in other words, left, right, up, or down). The C will appear as one of three colors (red, green, or blue) and vary in intensity. When each C is presented you will have eight seconds to make your response before the test continues on to present the next letter. This test is adaptive, which means that the test will adjust the difficulty level of the intensity based on your responses to estimate the point at which the letter is just barely noticeable. This test measures your threshold, so don't expect to get all of them correct. If you're ever unsure of the correct answer, just try to make your best guess. When you're ready to begin, please press the bottom button to cycle through the prompts."

A.2 LCST – Contrast Sensitivity

The old AVT instructions were the same as in the old OCCT (see A.1).

The new AVT instructions were as follows: "This test will be examining your contrast sensitivity. In the center of the screen you will see the letter C. The C will take one of four orientations: normal letter C, rotated downward, rotated left, or rotated up. Using the round buttons on the game controller, press the button that corresponds with the direction of the gap in the C (in other words, left, right, up, or down). The C will appear as three different sizes (large, medium, and small) and in various shades of gray. Prior to the C appearing on the screen, four bars in the shape of a box will appear, indicating the size of the next letter. When each C is presented you will have eight seconds to make your response before the test continues on to present the next letter. This test is adaptive, which means that the test will adjust the difficulty level of the visibility based on your responses to estimate the point at which the letter is just barely noticeable. This test measures your threshold, so don't expect to get all of them correct. If you're ever unsure of the correct answer, just try to make your best guess. When you're ready to begin, please press the bottom button to cycle through the prompts.

A.3 LCST - Acuity

The old AVT instructions were the same as in the old OCCT (see A.1).

The new AVT instructions were as follows: "This test will be measuring your visual acuity. In the center of the screen you will see the letter C. The C will take one of four orientations: normal letter C, rotated downward, rotated left, or rotated up. Using the round buttons on the game controller, press the button that corresponds with the direction of the gap in the C (in other words, left, right, up, or down). The C will appear as a solid black color and will vary in size. When each C is presented you will have eight seconds to make your response before the test continues on to present the next letter. This test is adaptive, which means that the test will adjust the size of the letter based on your responses to estimate the point at which the letter is just barely large enough to identify the orientation. This test measures your threshold, so don't expect to get all of them correct. If you're ever unsure of the correct answer, just try to make your best guess. When you're ready to begin, please press the bottom button to cycle through the prompts."

A.4 Dual Ring Stereo

The old AVT instructions were as follows: "On the center of the screen you will see two concentric circles, one inside the other, like a bullseye. The circles will be grey in color. When the test starts, you will indicate whether the smaller inner circle is popping out at you, or in other words, out in front of the larger outer circle, or whether the smaller inner circle is going back into the screen, or back behind the larger outer circle. If the inner circle looks like its popping out towards you, press the green 'A' button on the control (the one closest to you). If the inner circle looks like it's going back into the screen, press the yellow 'Y' button (the one furthest from you). We will begin with a short training round. This is just practice so no pressure."

The new AVT instructions were as follows: “On the center of the screen you will see two concentric rings, one inside the other. The rings will be gray in color. Your task is to determine whether the inner ring is popping out towards you or going back into the screen. If the smaller, inner ring looks like it is popping out towards you, or in other words, out in front of the larger outer ring, press the green bottom button. If the smaller, inner ring looks like it is going back into the screen, or in other words, behind the outer ring, press the yellow top button. The stimuli will appear on the screen for eight seconds before disappearing. You will have an unlimited amount of time to respond, and the test will not move on to the next trial until you respond, so if you are ever unsure of the correct answer just try your best to guess. There are three examples above. The left example shows what it will look like when the smaller ring is popping out at you. The right example shows what it will look like when the smaller ring is going back into the screen. The middle example is a randomized demo that you can use to practice. When you feel comfortable with the test and are ready to begin, please press the start button on the controller.”

A.5 Stereo Search

The stereo search test was added to the new AVT, so there are no old AVT instructions.

The new AVT instructions were as follows: “For this test, you will be presented with a plate containing four rings as shown in the examples above. While the rings may appear to be various distances from the screen, one of the rings should appear to be nearer to you than the rest. Identify which ring appears to be nearer to you and use the rounded buttons on the gamepad to indicate which position that ring is in. The stimuli will appear on the screen for eight seconds before disappearing. You will have an unlimited amount of time to respond, and the test will not move on to the next trial until you respond, so if you are ever unsure of the correct answer just try your best to guess. When you are ready to continue to the test, please press the bottom button on the controller.”

A.6 Fusion Range

The old AVT instructions were as follows: “On the center of the screen you should see a white circle. When the experiment starts, that circle is going to slowly pull apart into two circles. Your job is to try and keep that circle fused as one circle for as long as you can by focusing on it. You may feel your eyes pulling or straining as the circle tries to break apart, this is normal. Continue focusing on the circle and fighting that straining feeling to keep the circle together as long as you can. When you can no longer keep the circle fused and it has definitely broken apart into two circles, press the ‘A’ button on the controller. Once you press the ‘A’ button, the circles are going to slowly start coming back together again. Your job at this point will be to try and help them along and get them back to one circle as quickly as possible. When the circles are back to one, press the ‘A’ button on the controller again. The breaking apart and coming back together of the circle is one trial. There will be 8 trials. The first 4 will be horizontal and the last 4 will be vertical. After each trial it will bring you back to this screen where you can rest your eyes if you need to and then press the ‘A’ button to continue with the next trial.”

“A few things to keep in mind while doing this task… If you happen to blink or look away and the circle breaks, but you are quickly able to refocus and get the circle back to one, that is ok and will still count as the circle being together. Also, on some conditions you may notice a slight ghosting effect in which you might see a very light shadow of the circles moving apart while still simultaneously seeing the bright fused circle in the center of the screen. If this happens just try your best to ignore it and remain focused on the center bright fused circle.

The new AVT instructions were as follows:

[Trials 1 and 2]

"On the center of the screen you will see a white ring. When the experiment starts, the ring should appear to move away from you in depth (i.e., into the screen). Do your best to actively track the ring in depth by following it with your eyes (i.e., uncross your eyes) until the single ring splits into two separate rings. You may feel your eyes pulling or straining as you perform this test. Although this may be uncomfortable, try to maintain fusion, or track the ring in depth as long as you can. For example, hold your finger in front of your face and focus on it. Now, slowly move your finger farther out in front of you while continuing to focus on it. This is what it will feel like to look at the ring as it's moving away from you. When you can no longer keep the ring fused and it has broken completely apart into two separate rings, press the bottom button on the controller. Once you press the bottom button, the rings will slowly start coming back together again. Make an effort to adjust, or uncross your eyes, until you once again see a single ring. The ring will look like it's far into the screen and moving towards you. As soon as you once again perceive a single ring, press the bottom button. If you happen to blink or look away and the ring breaks apart, but you are quickly able to refocus and get the ring back to one that is okay and will still count as the ring being together. Also, you may notice a slight ghosting effect in which you might see a very light shadow of the rings moving apart while still simultaneously seeing the bright fused ring in the center of the screen. If this happens, just try your best to ignore it and remain focused on the bright fused center ring. You may start the experiment now by pressing the bottom button."

[Trials 3 and 4:]

"For the next two fusion range conditions, you should see a single ring that appears to move towards you in depth. Do your best to actively track the ring in depth by adjusting your eyes (i.e., cross your eyes) until the single ring completely splits apart into two separate rings. For an example, hold your finger out in front of your face and focus on it. Now slowly move your finger towards your nose while continuing to focus on it. This is what it will feel like to look at the circle as it's moving away from you. When you can no longer keep the ring fused and it has completely broken apart into two rings, press the bottom button on the controller. Once you press the bottom button, the rings will slowly start to come back together again. Make an effort to adjust, or cross your eyes, until you once again see a single ring. The ring will look like it's in front of the screen and moving away from you. As soon as you once again perceive a single ring, press the bottom button. You may start the experiment now by pressing the bottom button."

[Trials 5-8:]

"For the next four fusion range conditions, you should once again see a single ring. Do your best to adjust your eyes and actively maintain focus on the ring until the single ring splits into two separate rings. This time, the rings will appear vertically separated. When you can no longer keep the ring fused and it has completely broken apart into two rings, press the bottom button on the controller. Once you press the bottom button, the rings will slowly start to come back together again. Make an effort to adjust your eyes until you once again see a single ring. As soon as you once again perceive a single ring, press the bottom button. You may start the experiment now by pressing the bottom button."

[Actual Experiment:]
"Now that we're finished with the training, the experiment will run through each of the conditions that you just practiced. After each trial, the experiment will bring you back to this screen where you should briefly rest your eyes. When you're ready to start the next trial, press the bottom button to begin the next trial. You may start the experiment now by pressing the bottom button.

A.7 Motion Perception

The old AVT instructions were as follows: "As you can see, there is a collection of dots that are either moving clockwise or counterclockwise. Your job is to identify the directions that the dots are moving, as a whole, and respond appropriately. If the collection of dots is moving clockwise, hit the 'right' arrow key on the keyboard. If the collection of dots is moving counterclockwise, hit the 'left' arrow key on the keyboard. For this next session, the dots are now either expanding outward or contracting in towards the center. If the dots are expanding outwards, like an exploding firework, press the 'up' arrow key on the keyboard. If the dots are contracting in towards the center, press the 'down' arrow key on the keyboard."

The new AVT instructions were as follows: "This test will be measuring your motion perception, and you will be using the arrow keys on the keyboard to respond. This test has two parts. For the first part, you will see a collection of dots in the shape of a large circle. As a whole, this group of dots will either be moving clockwise or counterclockwise. If the dots are moving clockwise, press the right arrow key. If the dots are moving counterclockwise, press the left arrow key. For the second part, the collection of dots will either appear to be expending outward, like a firework, or expending inward towards the center of the circle. If the dots are expanding outward, press the up arrow. If the dots are contracting inward, press the down arrow. During the test you should look at the center of the circle, but also attend to the whole field of dots. Don't narrowly focus on only one area of the circle or you may not be able to determine which direction, as a whole, the collection of dots is moving. This test is adaptive, which means that, based on your responses, the test will adjust the proportion of dots moving in a unified direction versus moving randomly, to the point where the direction of motion is just barely noticeable. For each trial, the dots will only appear briefly, but you will have as much time as you want to respond."

Appendix B – AVT Histograms

B.1 LCST

B.1.1 Contrast Sensitivity

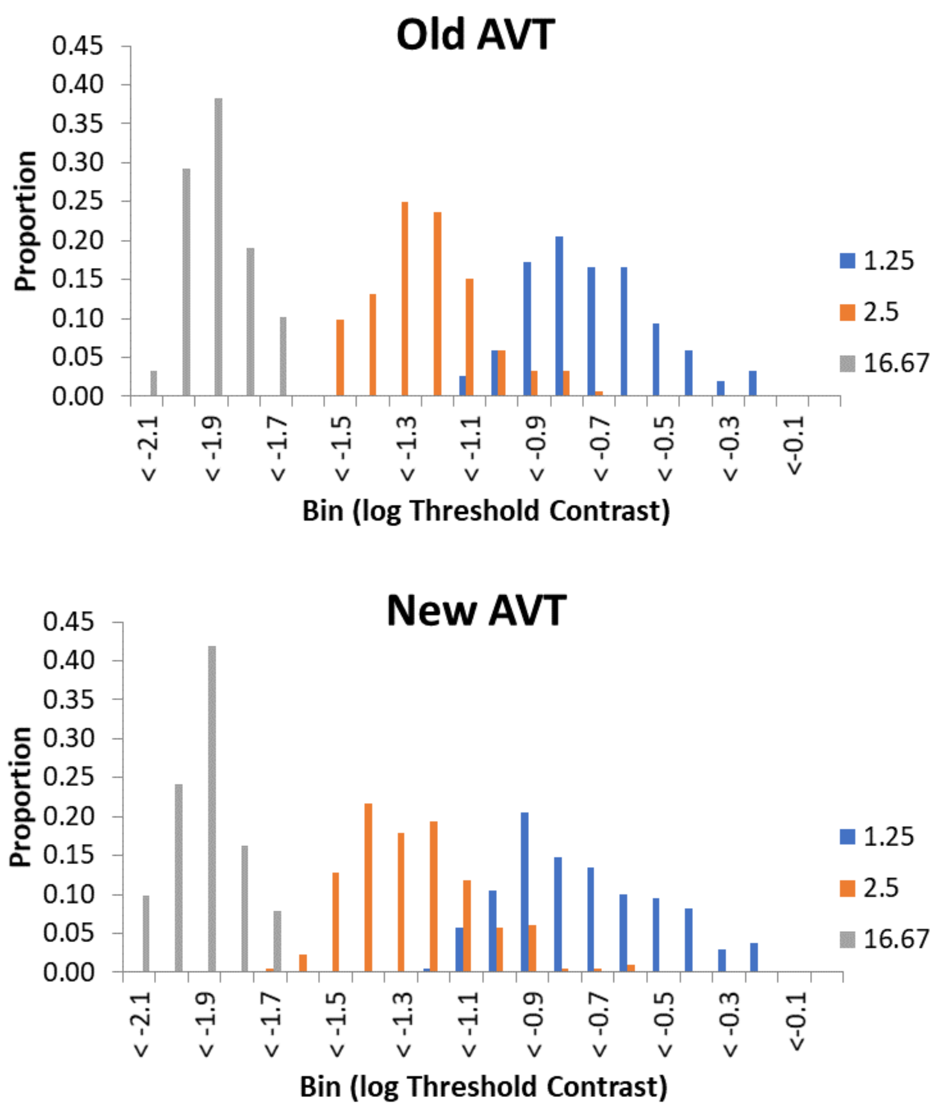


Figure B-1. Contrast sensitivity test histograms. Top: old AVT. Bottom: new AVT.

B.1.2 Acuity

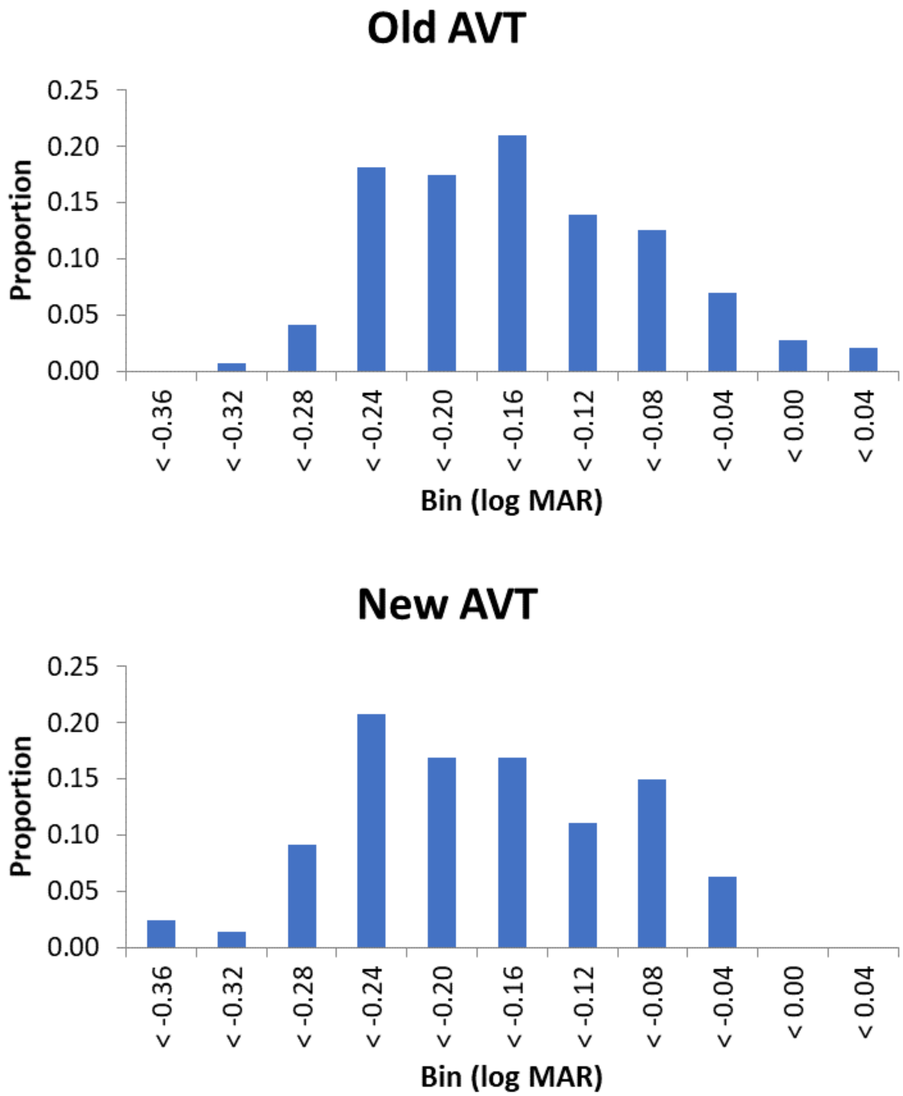


Figure B-2. 4M acuity test histograms. Top: old AVT. Bottom: new AVT.

B.2. OCCT

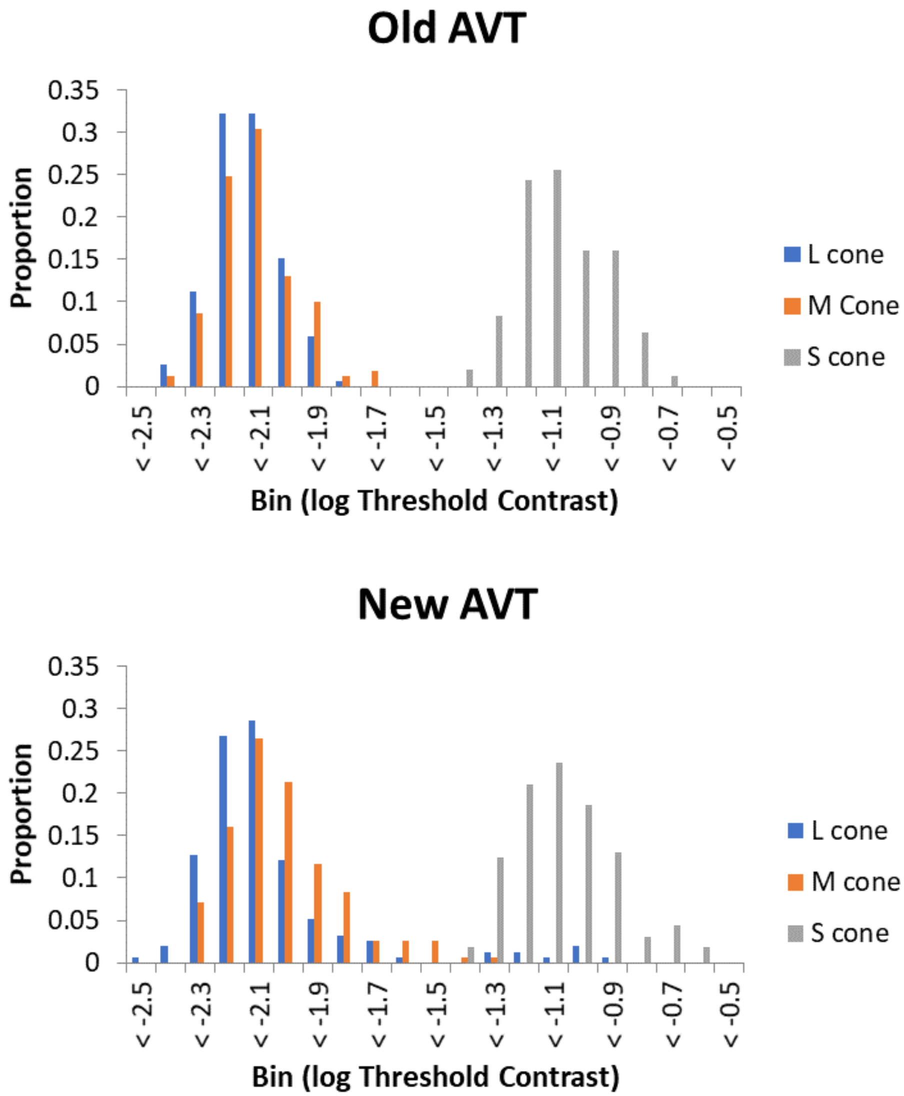


Figure B-3. OCCT test histograms. Top: old AVT. Bottom: new AVT.

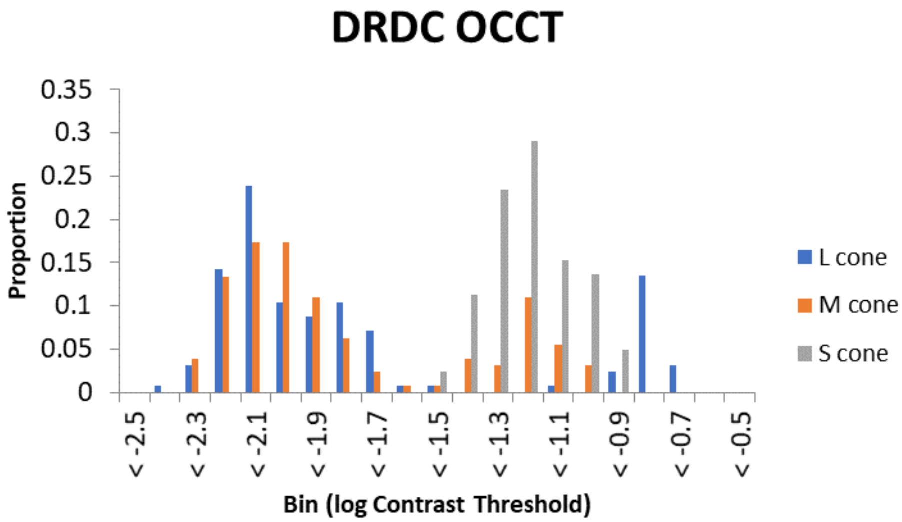


Figure B-4. OCCT test histogram from data collected at the DRDC.

B.3. Dual Ring Stereo Acuity

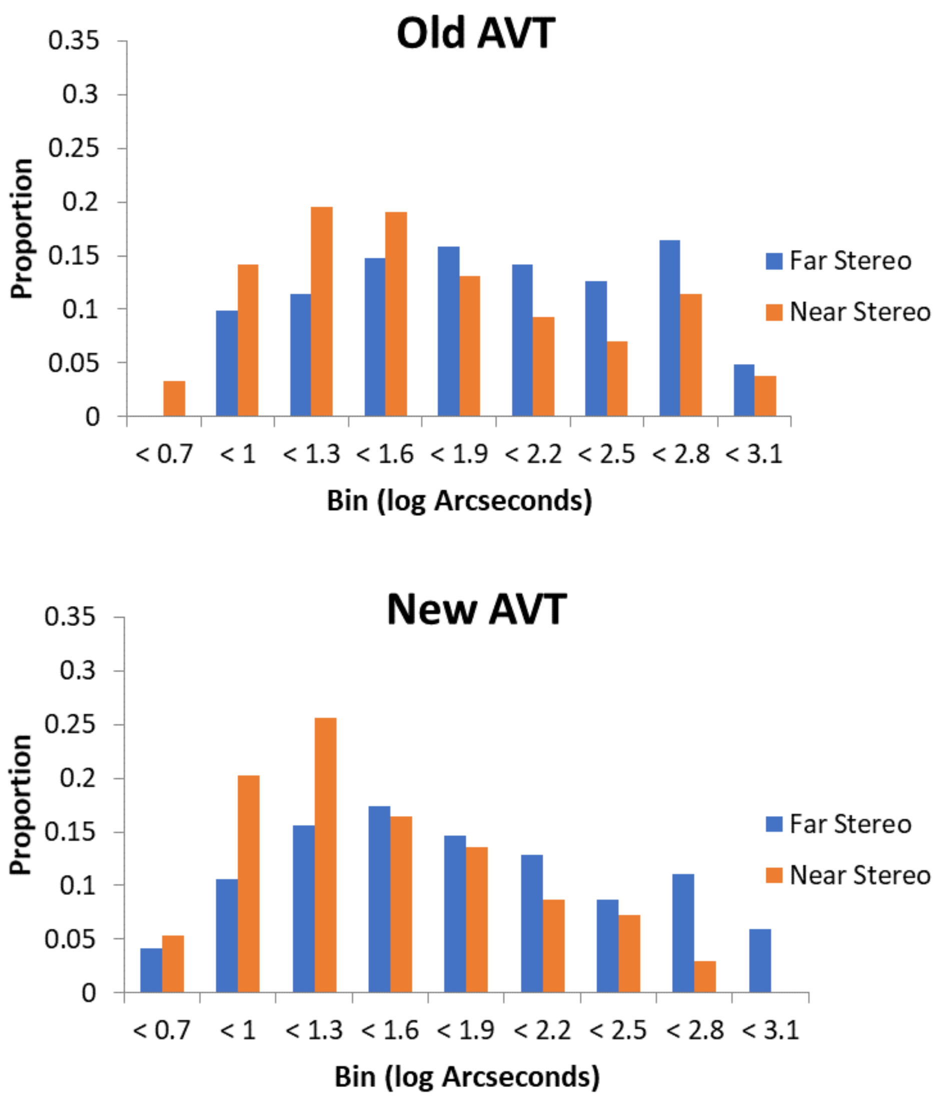


Figure B-5. Dual ring test histograms. Top: old AVT. Bottom: new AVT.

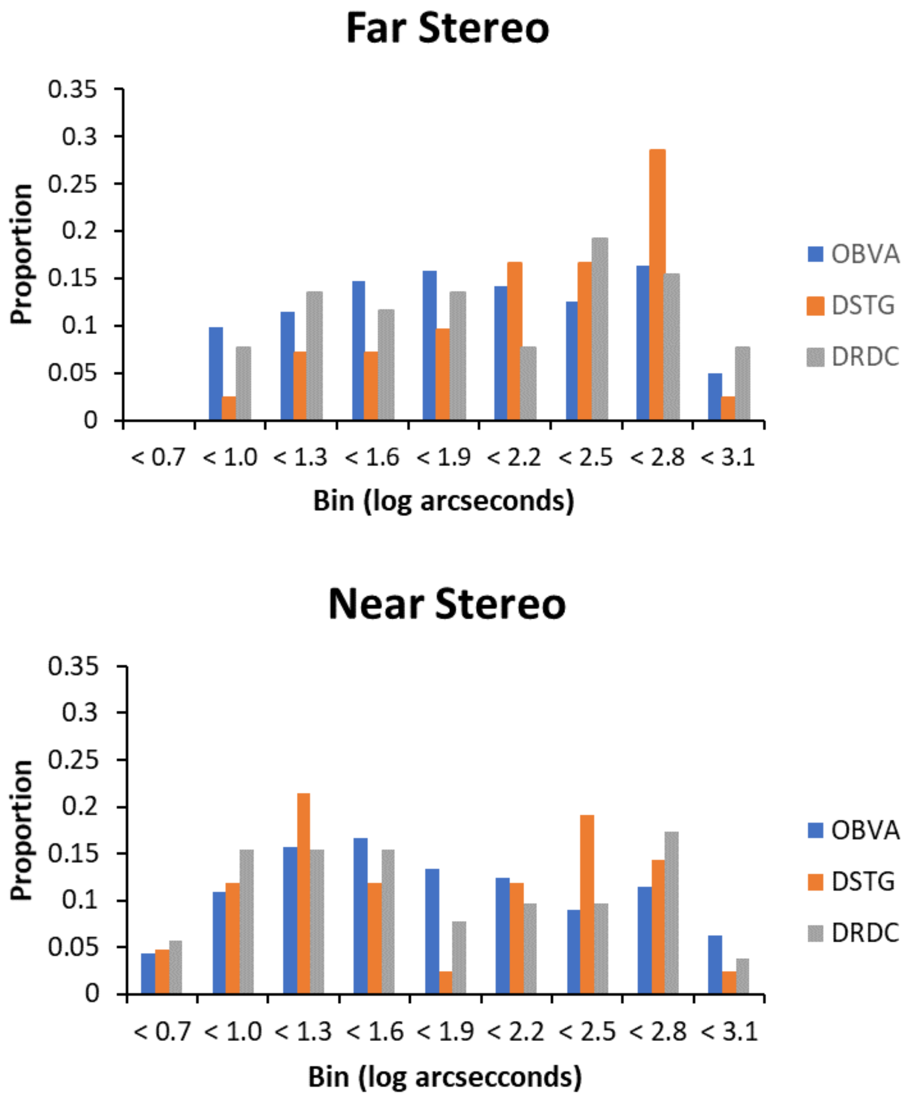


Figure B-6. Dual ring test histograms with data collected in different locations. Top: far stereo. Bottom: near stereo.

B.4. Stereo Search

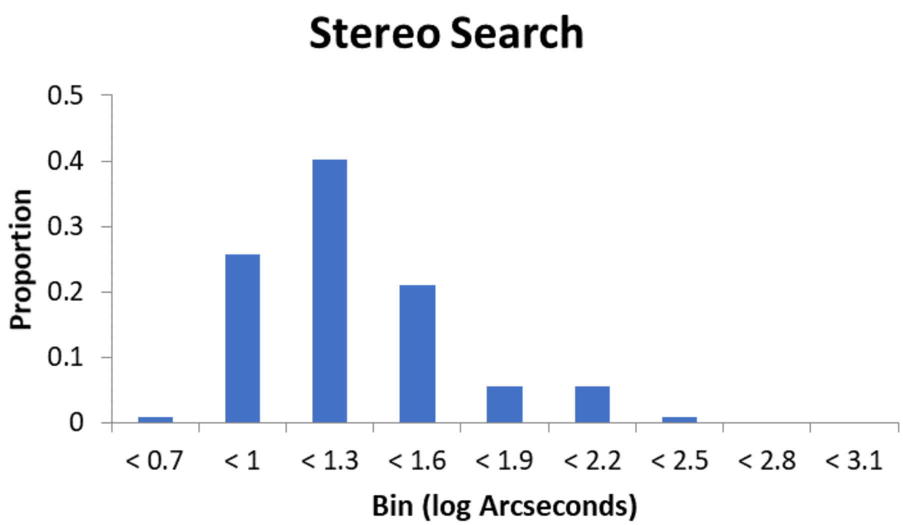
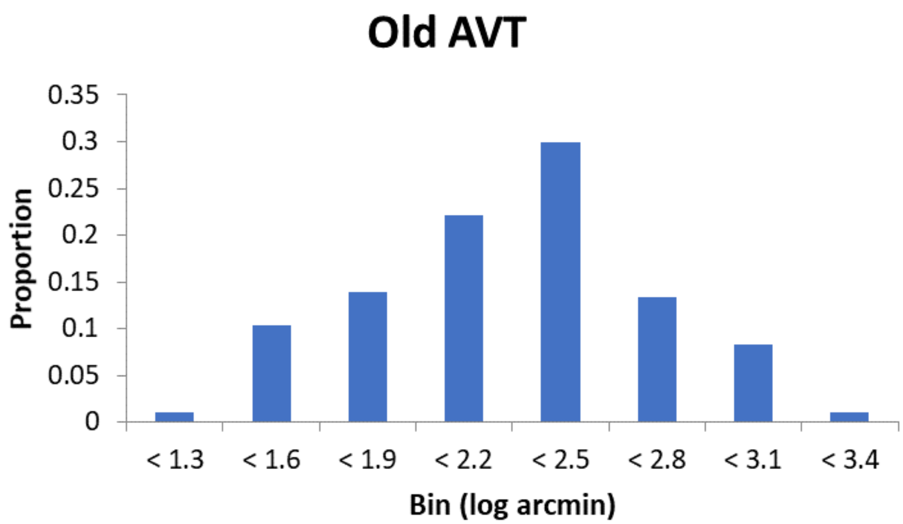
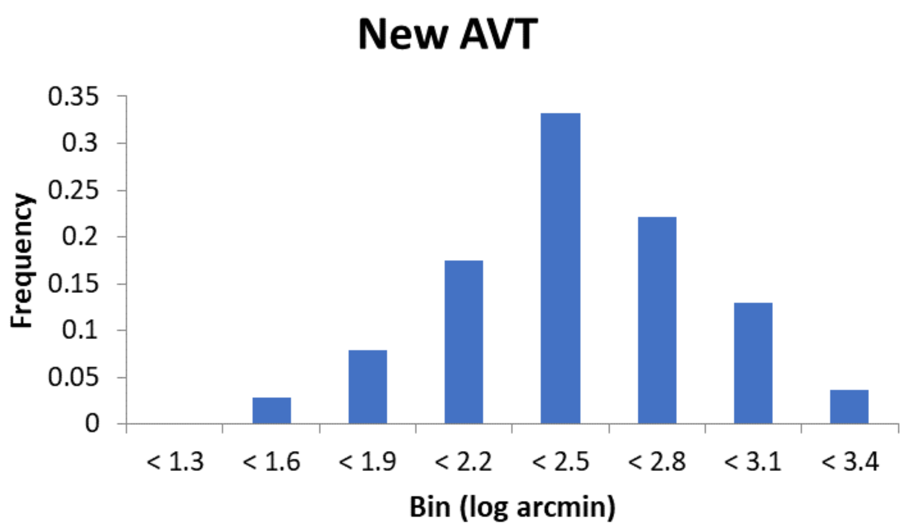


Figure B-7. Stereo Search test histogram.

B.5. Fusion Range

B.5.1. Horizontal Fusion Range





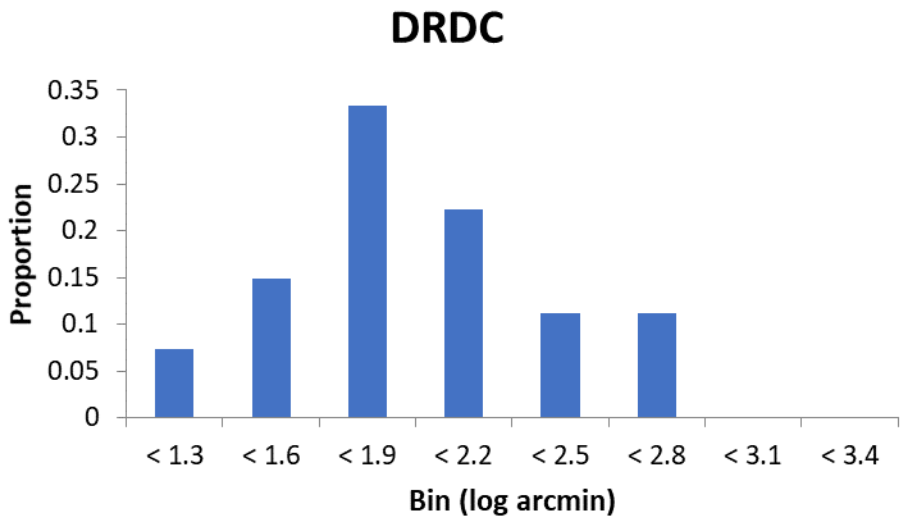
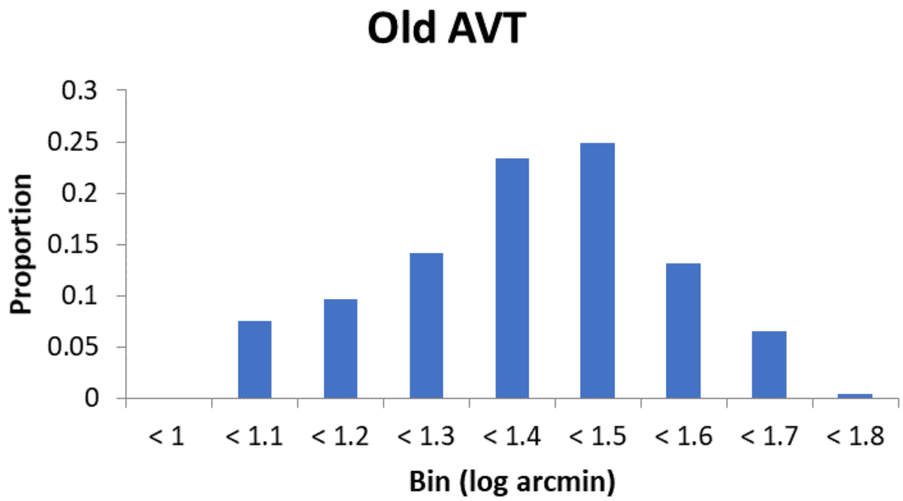
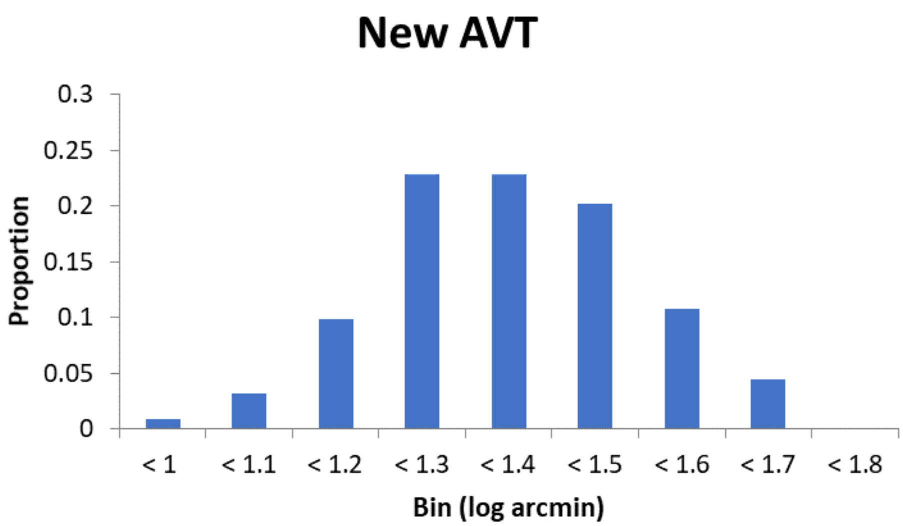


Figure B-8. Horizontal fusion range test histograms. Top: old AVT; Middle: new AVT; Bottom: DRDC old AVT

B.5.2 Vertical Fusion Range





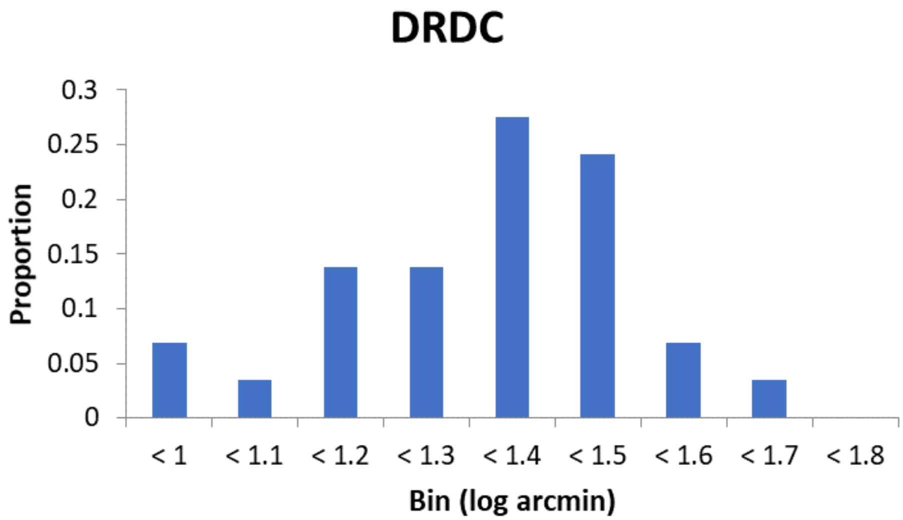


Figure B-9. Vertical fusion range test histograms. Top: old AVT; Middle: new AVT; Bottom: DRDC old AVT

B.6. Motion

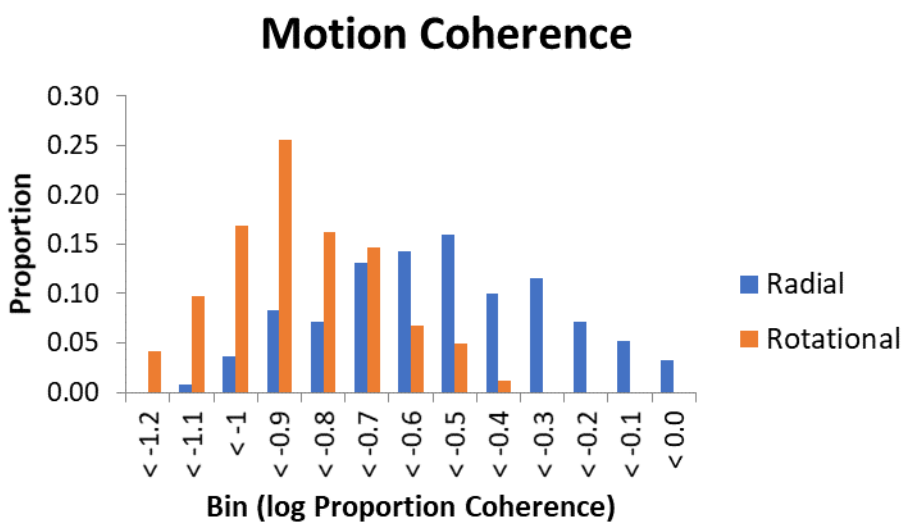


Figure B-10. Motion test histogram.