Federal Aviation Administration

DOT/FAA/AM-13/16
Office of Aerospace Medicine
Washington, DC 20591

Evaluation of the Stereo Optical OPTEC® 5000 for Aeromedical Color Vision Screening

Nelda J. Milburn
Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

Carrie A. Roberts
Deborah L. Perry
Xyant Technology, Inc.
Oklahoma City, OK 73125

August 2013

Final Report

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-13/16		

4. Title and Subtitle	5. Report Date
Evaluation of the Stereo Optical OPTEC® 5000 for Aeromedical Color Vision Screening	August 2013
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Milburn NJ,[1] Roberts CA,[2] Perry DL[2]	

9. Performing Organization Name and Address		10. Work Unit No. (TRAIS)
[1]FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125	[2]Xyant Technology, Inc. Oklahoma City, OK 73125	11. Contract or Grant No.

12. Sponsoring Agency name and Address	13. Type of Report and Period Covered
Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	14. Sponsoring Agency Code

15. Supplemental Notes

Work was accomplished under approved task AM-521.

16. Abstract

Screening tests are valued for their ability to detect the presence (test sensitivity) and the absence (test specificity) of a disease or a specific condition such as color vision deficiencies (CVDs). From an aviation safety standpoint, it is important to identify those with CVDs because of their potential for accidents if they misinterpret vital color-coded information; whereas, failing an airman with normal color vision has other consequences to the Federal Aviation Administration (FAA) such as the expense of secondary screening. If a screening test has low specificity, it can have a high false positive rate—meaning that individuals are falsely noted as having the condition being screened. So, from the airman's standpoint, especially if the airman has normal color vision (NCV), specificity is very important. The FAA has a color vision standard for airmen and air traffic controllers because of the occupations' high reliance on color-coded information.

Stereo Optical Company, Inc. requested a review of their model 5000 multifunction screener for aeromedical use, and the FAA found that instrument failed 50% of those with NCV. The manufacturer made some modifications and requested a re-evaluation. The validity of the original and modified versions of the OPTEC® 5000 (called V1 and V2, respectively in this paper) was examined in two experiments.

Experiment 1 involved 29 NCV and 31 CVD subjects that were administered the OPTEC® 2000 and the V1. Experiment 2 examined the validity of the original and modified instruments by comparing test outcome on both versions to a diagnostic test using 50 NCV and 51 CVD subjects.

In Experiment 1, the V1 failed 41.3% of the NCV participants. In Experiment 2, V1 failed 28%, and V2 failed 32% of the NCV subjects. V1 passed 16% of the subjects diagnosed with CVDs and V2 erroneously passed 12%. It is possible that the light-emitting diode strip, used in the OPTEC® 5000, altered the color perception of the pseudo-isochromatic test plates and that substituting that light source for one with better color rendering could restore the test's sensitivity and specificity rates. The specificity of the OPTEC®5000, in its original and modified states, are unacceptably low; and neither should be used for aeromedical color vision screening because of their low agreement with a diagnostic test ($K_{(101)}$=.564 and .563, respectively).

17. Key Words	18. Distribution Statement
Color Vision Deficiency, Color Vision, Color Vision Screening Tests, OPTEC 5000	Document is available to the public through the Internet: www.faa.gov/go/oamtechreports

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	18	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

CONTENTS

Evaluation of the Stereo Optical OPTEC® 5000 for Aeromedical Color Vision Screening

Introduction ---------- 1
Methods ---------- 1
Materials ---------- 2
Subjects ---------- 2
Procedure ---------- 3
Results and Discussion ---------- 3
Experiment 1: Comparison of OPTEC® 2000 with OPTEC® 5000V1 ---------- 3
Experiment 2: Comparison of OPTEC® 5000V1 to OPTEC® 5000V2 ---------- 6
Summary of Experiments 1 and 2 ---------- 8
Conclusions ---------- 10
References ---------- 11
Appendix A ---------- A1

Evaluation of the Stereo Optical OPTEC® 5000 for Aeromedical Color Vision Screening

INTRODUCTION

Screening tests are valued for their ability to detect the presence and the absence of a disease or a specific condition such as color vision deficiencies. Screening tests are rated by their ability to detect a condition known as a test's sensitivity. Likewise the test's ability to evaluate the absence of a disease is equally important. For example, cancer patients are eager to receive the good news that there is no evidence of cancer. A screening test's accuracy for detecting the absence of a condition is called its specificity.

From an aviation safety standpoint, it is important to identify those with color vision deficiencies (CVDs) because of their potential for accidents if they misinterpret vital color-coded information such as a precision approach path indicator (PAPI) light; whereas, failing an airman with normal color vision has other consequences to the Federal Aviation Administration (FAA) such as the expense of secondary screening. If a screening test has low specificity, it can have a high false positive rate meaning that individuals are falsely noted as having the condition being screened. In terms of color vision deficiency, a test with a high false positive rate is a test that denotes a normal color vision individual as having a deficiency, and that is comparable to diagnosing a well person with a disease. What that means to a pilot applicant is failing a critical aeromedical criterion unnecessarily and unfairly. Therefore, it is important that a color vision screening test has both high sensitivity and high specificity.

Validating a screening test and measuring the sensitivity and specificity requires a repeated-measures design, accomplished by obtaining performance data on both the screening test and a diagnostic test with sufficient subjects in both outcome categories (e.g., those with normal color vision and those with color vision deficiencies). Calculation of sensitivity and specificity for a screening test is determined by comparing outcome on a screening test to outcome on a criterion measure, which for color vision includes such diagnostic tests as the Nagel Type 1 anomaloscope, the Oculus anomaloscope, the Colour Assessment and Diagnoses (CAD) test, and a few others. Some may ask, "Why not simply use a diagnostic test exclusively to provide the definitive conclusion regarding one's color vision status?" The answer is that, typically, screening tests are valued, used, or preferred over diagnostic tests because they possess one or more of the following attributes: They are quicker to administer, require less skill to administer, are less expensive, more accessible, or the screening test has some additional functions such as measuring visual acuity or contrast acuity. For example, the Nagel anomaloscope is considered the gold standard for diagnosing color vision deficiencies of the red-green type; however, it requires a highly-skilled test administrator, takes about 20 to 30 min to administer, and is not readily available for purchase. In contrast to the Nagel anomaloscope, most pseudo-isochromatic plate (PIP) tests are about 1/100th of the anomaloscope's price and some have kappa values greater than .9, meaning that they agree with the outcome of a diagnostic instrument about 90% of the time. Training to administer PIP tests is minimal and screening takes about 5 min, all factors that make them an attractive alternative to the diagnostic tool. Several PIP tests (e.g., the Ishihara, Dvorine, Waggoner, and Richmond® HRR) are commercially available, and some multifunction screening tests, such as the Stereo Optical OPTEC® 2000, the Titmus® i400, and others all make use of integrated PIP plates for measuring color vision in addition to other vision screening tests.

The FAA Civil Aerospace Medical Institute, Aerospace Human Factors Division examined the validity of the OPTEC® 2000, along with all other currently available color vision screening tests (Mertens & Milburn, 1993); and as a result, the OPTEC® 2000 appeared on the FAA's list of accepted color vision screening tests, the Guide to Aviation Medical Examiners (FAA, 1992). The FAA currently maintains that list on-line (FAA, 2013).

The OPTEC® 5000 was developed to replace the OPTEC® 2000; however, when the Civil Aerospace Medical Institute's, Vision Research group evaluated Stereo Optical's newer model, the OPTEC® 5000, it did not perform as well as its predecessor for color vision screening, "…it failed 50% of the color normal subjects in the study" (Nakagawara, Montgomery, & Wood, 2009, p.1). Unfortunately, modifications and updates that were intended to improve screening performance actually degraded the new version's specificity, its ability to dismiss the presence of a color vision deficiency. As a result, the OPTEC® 5000 was not added to the FAA's list of approved tests for color vision screening. Consequently, Stereo Optical made some additional modifications to the OPTEC® 5000 in an attempt to create a valid color vision screening test and asked FAA personnel to re-evaluate the OPTEC® 5000.

The purpose of this report was to evaluate the validity of the modified OPTEC® 5000 for screening color vision, and to do so, OPTEC® 5000 test outcome was compared to diagnosis (normal color vision vs. color vision deficiency) on the CAD test.

METHODS

Prior approval for all procedures and use of human subjects was obtained from the FAA Institutional Review Board. Informed consent was obtained prior to participation, and subjects were free to withdraw from the project without consequence at any time.

Research reported in this paper was conducted under the Flight Deck Program Directive / Level of Effort Agreement between the Federal Aviation Administration Headquarters and the Aerospace Human Factors Division of the Civil Aerospace Medical Institute and was sponsored by Office of Aerospace Medicine and supported through the FAA NextGen Human Factors Division.

Materials

Colour Assessment and Diagnosis (CAD) Test. The Colour Assessment and Diagnosis (CAD) Test (distributed by City Occupational, Ltd., London) is a computerized color vision test that screens for normal color vision, quantifies loss of chromatic sensitivity, and classifies individuals by type and degree of color vision deficiency. The full, definitive CAD test takes about 15 min to complete. The participant's task is to indicate the direction of movement of a colored square target on the dynamic checkerboard background via a response pad that employs a four-alternative, forced-choice procedure with each of the four buttons corresponding to the four diagonal directions of movement. The very large number of trials prevents examinees from learning responses, which is possible on the limited trials of pseudoisochromatic plate tests. As an added benefit, the CAD test plots the individual's chromatic discrimination sensitivity in the Commission Internationale de l'Eclairage (CIE) 1931 color space and provides both red/green and yellow/blue thresholds relative to the standard normal observer and reports those threshold values in standard normal units (SNU), such that a threshold value of an individual indicates the normed value for the standard normal observer. No color naming is involved. The viewing distance from the 17-inch ViewSonic E70fSB CRT monitor is 140 cm (~55 inches). The illumination falling on the desktop in the testing room averaged about 10 to 15 lux.

Signal Light Gun Test (SLGT). The signal light gun test used the Model 901 (distributed by ATS Aerospace, Inc., Canada) signal light gun. The SLGT has a unique distinction, in that it is the actual instrument used by air traffic control specialists to communicate with pilots, but is also the same instrument used to determine whether a pilot receives a "waiver" for color vision as a Statement of Demonstrated Ability (SODA). If a pilot applicant fails an initial color vision screening test administered by an aviation medical examiner (AME), then applicants for a first- or second-class medical certificate are required to take and pass an Operational Color Vision Test (OCVT) and a color vision Medical Flight Test (MFT). Applicants for a third-class medical certificate need only to take and pass the OCVT. The OCVT has two components, the SLGT and demonstration of the ability to correctly read and interpret colors on aeronautical charts (Code of Federal Regulations, 2013). The SLGT is presented at two distances, a near distance of 1,000 ft (304.8 m) and a far distance of 1,500 ft (457.2 m). When the SLGT is given to pilot-applicants by FAA Flight Standards District Office aviation safety inspectors, testing at the near distance is always first. However, as part of a separate study to determine whether continued testing at both distances is necessary, the ordering of the near and far distances alternated throughout the experimental trials. The colors within each distance test site were given in the same order for all participants. In actual pilot applicant testing, examinees receive six trials at each distance with the three colors randomly ordered, with each color presented at least once at each distance. Each participant was asked to write the name of the color presented on the answer sheet provided, for each trial. The pass criterion was zero errors among the 12 trials.

Stereo Optical® Vision Testers. Two Stereo Optical® models were used, a model 2000 (OPTEC® 2000) and a model 5000 (OPTEC® 5000); spec sheets for both models are available from the manufacturer. Both instruments are considered multifunction visual screening instruments; however, only the color vision screening test (Slide 2000-010 "FAR" Color Perception) was evaluated in this experiment. The color vision screening test consists of a single pseudo-isochromatic plate containing six trials (called A through F), with all trials being visible at once. Three identical copies of the pseudo-isochromatic plate were used—one residing in the OPTEC® 2000 with its incandescent light source of four 7-watt bulbs, part # 2000-226 (x=.326, y=.261), and two plates, each residing in a different slot of the OPTEC® 5000, which will be referred to as the *original plate* (OPTEC® 5000V1) and the *modified plate* with a manufacturer-applied, orange film (Rosco filter #3441 -full straw) covering the plate (OPTEC® 5000V2). The OPTEC® 5000 apparatus uses a light-emitting diode (LED) strip (lighting systems part # 520-49) containing four LEDs (x=.414, y=.384, ~3200K) to illuminate the test slides. The OPTEC® 5000 makes use of a knob to change presentation slides for the various vision tests (visual acuity, color perception, lateral/vertical phoria, fusion, muscle balance, stereo depth, and tumbling "E" perception) that reside in separate slots.

Subjects

Data for two separate studies are presented: a study conducted in 2010 with 60 subjects that responded to both the OPTEC® 2000 and the OPTEC® 5000V1 will be referred to as Experiment 1. A separate, follow-on study conducted in 2011 with 101 subjects, comparing performance on the OPTEC® 5000V1 and the OPTEC® 5000V2, will be called Experiment 2. One difference between the two studies was the age restrictions for the subjects. The subjects of Experiment 1 were intended to be reflective of pilots for a study involving airport lighting, and their age ranged between 18 and 58 years. Subjects of Experiment 2 were recruited specifically for a larger study meant to relate to air traffic control applicants, so, the subjects were restricted to those18-33 years of age. All subjects of both studies were screened for visual acuity for both near and far vision using the Bausch and Lomb Orthorater (Bausch and Lomb, Rochester, NY), and subjects met a criterion of at least 20/30 (with correction, if necessary). In both studies, color vision classification was determined by the Colour Assessment and Diagnosis (CAD) test, and participants were categorized by color vision type (protan, deutan, or tritan). Readers are directed to Barbur, Rodriguez-Carmona, and Harlow (2006) for an in-depth description of the CAD test and Barbur, Cole, and Plant (1997) for an explanation of the various types of color vision deficiencies, and Sharpe, Stockman, Jagle, and Nathans (1999) for the prevalence within the population of each type of deficiency.

Experiment 1 volunteers were from the Troy, New York, commuting area, recruited and paid by a contractor. Participants were 29 individuals with normal color vision (NCV) and 31 with color vision deficiencies (CVD) classified as follows: 12

protans, 15 deutans, 1 tritan, and 3 subjects evidencing both red-green and yellow-blue weaknesses. NCV participants were 5 females and 24 males with a mean age of 27.2 years, SD of 7.9 years and CVD participants were 29 males and 2 females with a mean age of 32 years, and SD of 11.3 years. The minimum age was 18 and the maximum was 58 years.

Participants of Experiment 2 were 50 NCV and 51 CVD individuals classified as follows: 12 protans, 29 deutans, 2 tritans, and 8 subjects evidencing both red-green and yellow-blue weaknesses. NCV participants were 26 females and 24 males with a mean age of 23.9 years, SD of 3.5 years; CVD participants were 36 males and 14 females with a mean age of 23.8 years and SD of 3.8 years. The minimum age was 18 and the maximum was 33. Study volunteers were from the Oklahoma City, Oklahoma, commuting area that were recruited and paid by a contractor.

Procedure

In both studies, participants were asked to complete several color vision screening, diagnostic, and occupational (color-naming or color matching) tests. Order of presentation of the Stereo Optical equipment was controlled such that about half received the OPTEC® 2000 before the OPTEC® 5000V1 in Experiment 1, and about half received the OPTEC® 5000V2 before the OPTEC® 5000V1 in Experiment 2. The participant's task was simply to record the numbers seen on each trial (labeled A through F) and to write "Blank" if they did not see any numbers for a trial. A test administrator closely monitored each test; and, in the case of the OPTEC® 5000, the test administrator adjusted the knob to ensure that the proper test version was presented, carefully matching the test to the labeled answer sheet. As previously mentioned, both studies were part of a larger study involving several color vision screening tests, so order of presentation was controlled, and several other tests occurred between the two versions being studied and reported here.

RESULTS AND DISCUSSION

The results are generally arranged by experiment and cover test performance (a) by CAD type diagnosis, (b) by comparing test versions, (c) by contrasting each test trial (A-F) by version as a function of color vision category (comparing NCV to CVD), (d) by examining the relationship between the SLGT outcome and specificity rates, (e) by examining test validity, calculated via Kappa (a measure of agreement after accounting for chance) using the CAD NCV and CVD categories as the criterion, and finally, (f) by reporting test sensitivity and specificity for OPTEC® 2000, 5000V1, and 5000V2.

Experiment 1: Comparison of OPTEC® 2000 with OPTEC® 5000V1

Using a repeated-measures design, all 60 subjects responded to both the OPTEC® 2000 and the OPTEC® 5000V1, with about half of the subjects responding to the OPTEC® 2000 first. Because both instruments used the same pseudoisochromatic plate containing six items, the hypothesis was that performance would be essentially identical if all other factors remained the same. However, Table 1 shows inconsistent pass/fail outcome performance for 10 individuals (16.67%), resulting in a Kappa agreement score of .654. Table 2 was created to explore this inconsistency and shows that one Deutan CVD, the only Tritan CVD, and six NCV participants failed the OPTEC® 5000V1 but passed the OPTEC® 2000.

According to Table 2, the OPTEC® 2000 failed 8 (27.5%) of the 29 NCV participants, whereas the OPTEC® 5000V1 failed 12 (41.3%) of the NCV group. These findings are consistent with the previous study conducted by Nakagawara et al. (2009) that showed much better performance by the OPTEC® 2000 than the newer replacement model. Of course, the question is what caused the disparity? To investigate that dilemma,

Table 1. Crosstabulation of the Pass/Fail Outcome of the OPTEC®2000 by the OPTEC®5000V1

		OPTEC® 5000V1		
OPTEC® 2000				
		Fail	Pass	Total
	Fail	33	2	35
	Pass	8	17	25
	Total	41	19	60
Kappa = .654				

Table 2. Crosstabulation of the Pass/Fail Outcome of the OPTEC® 2000 by the OPTEC® 5000V1 by CAD Type Diagnosis

		OPTEC® 5000V1	
	OPTEC® 2000		
Normal	Fail	6	2
(n=29)	Pass	6	15
Protan	Fail	12	0
(n=12)	Pass	0	0
Deutan	Fail	14	0
(n=15)	Pass	1	0
Tritan	Fail	0	0
(n=1)	Pass	1	0
RG & YB	Fail	1	0
(n=3)	Pass	0	2

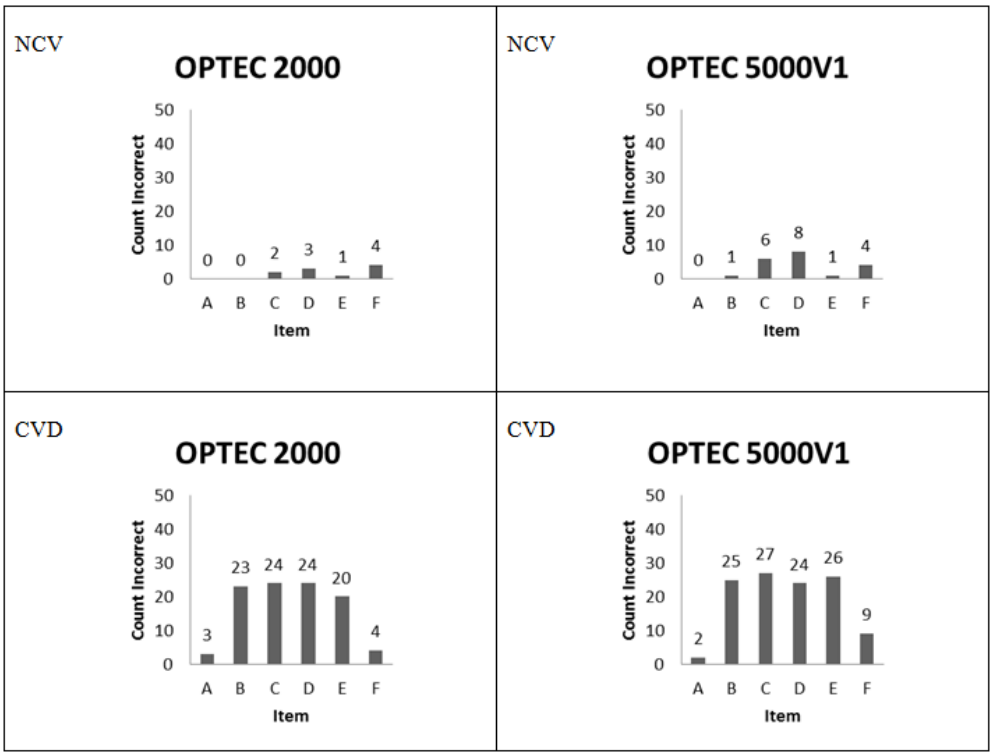

Figure 1. Count of the number of incorrect responses made by NCV (n=29) and CVD (n=31) participants in Experiment 1 comparing the OPTEC® 2000 and the OPTEC® 5000V1 by item (A-F).

correct/incorrect performance on individual items was compared between versions, and because previous findings indicated that NCV participants failed the 5000V1 model more often than the 2000 model, separate figures were constructed for convenient side-by-side comparison. Figure 1 shows a small increase in failures for items B, C, E, and a marked increase for item F for CVD participants and items C and D for those with NCV. Almost triple the number of NCV participants failed items C and D on the OPTEC® 5000V1 as failed the same items on the OPTEC® 2000, so the problem consolidates to "Why? What is unique to those items that would 'trip up' NCV participants?"

The appropriate approach to scientifically examine those items, as presented on each model, would be to make chromaticity measurements of the individual dots on the pseudo-isochromatic plates to look for variations between the plates (simply to rule that factor out as a potential cause) and then to make color rendering measurements of the light sources. Color rendering is defined as the "effect of an illuminant on the color appearance of objects by conscious or subconscious comparison with their color appearance under a reference illuminant" (CIE, 1987). The Commission Internationale de l'Eclairage (CIE) first proposed a Color Rendering Index (CRI) in 1964, updated it in 1974, and is a metric used to assess the ability of an artificial light source to render visible colors. If the artificial light source perfectly renders a color as well as the natural light source, an index of 1.0 is achieved. "The CRI has shortcomings in application, however, and its problems are pronounced when applied to newer lighting technologies, such as light-emitting diodes (LEDs)" (Davis & Ohno, 2009, p. 1412). Current CRI has been shown to incorrectly estimate the color rendering capabilities of LEDs (CIE 1995), and several alternative methods have been recommended by others (CIE 1995; Davis & Ohno, 2005, 2006, 2010; Ohno, 2004, 2005; Quintero, Sudria, Hunt, & Carreras, 2012).

Because the CRI is calculated as an average of 8 colors to indicate how closely the color appearance is under a light source compared to its appearance under natural daylight, the index is relevant as a broad interpretation of the light source. That averaging formula "makes it possible for a lamp to score quite well, even when it renders one or two colors very poorly" (Davis & Ohno, 2009, p. 1415). In general, that index is valuable, but for specific applications such as choosing an appropriate light source to enhance the appearance of red meat in a grocery store display, the index may not *tell the whole story*. For example, the color rendering may be good for 7 of those colors but poor for red; therefore that light source may make the meat appear brown and hence, less appealing to customers. That same light source may be a good choice for green leafy vegetables. "LEDs are at an increased risk of being affected by this problem, as their peaked spectra are more vulnerable to poor rendering in only certain areas of color space" (Davis & Ohno, 2009, p. 1415). Likewise, a LED may or may not provide good color rendering for all of the colors within a PIP test.

There are several viable methodologies for finding the cause for the increased failure of NCV participants on items C and D for the OPTEC® 5000V1—the approach *could* include measurements of all of the unique colors used in the test plates (to serve as a set of reflective samples), taken under natural light

and under the OPTEC® 5000V1 LED light source to calculate a color rendering index specific to this application. "Several proposed color rendering assessment methods share the basic procedure of the CRI: the appearance of a predetermined set of reflective samples when illuminated by the test source is compared to their appearance under a reference illuminant" (Davis & Ohno, 2009, p. 1416). Davis and Ohno (2009) have reported that some LEDs do not have a good color rendering of the color red, or sometimes, only specific areas of the color space. Knowing that, and how pseudo-isochromatic test plates make use of very subtle color differences, it is easy to understand that even small decrements in color rendering (perhaps only affecting one color that is used to form the numeral among distraction dots on a PIP) can have detrimental effects for color vision tests. Although these proposed methods may be the correct approach to *thoroughly* investigate the root cause for the increased errors that caused some NCV participants to fail, each strategy would require a sensitive spectroradiometer equipped with an appropriate LED sensor, a time-consuming investigation, and an investigation well beyond the scope of this report that should be reserved for color vision test manufacturers, lighting manufacturers, and the National Institute of Standards and Technology (NIST), rather than the FAA.

In lieu of that sophisticated approach, four researchers with normal color vision made side-by-side comparisons of the OPTEC® 2000 and the OPTEC® 5000V1 items C and D and explained the visual difference being less salient targets on the OPTEC® 5000V1, meaning that the hidden number was harder to distinguish from the background dots, essentially the same problem that CVDs experience with PIP tests.

Appendix A contains six tables that directly compare performance on the OPTEC® 2000 to the OPTEC® 5000V1 by color vision type classification for each item (A-F). The purpose of those tables was to explore whether certain types of deficiencies were more affected than others on specific items. Multiple, unequal groups with small sample sizes made most statistics, even for repeated measures, untenable choices. Therefore, the tables are simply presented without the usual, accompanying statistics for definitive results.

At the conclusion of Experiment 1, a representative from Stereo Optical contacted researchers at the FAA and submitted a prototype plate for evaluation. It was the original color vision plate covered with an orange film (Rosco filter #3441-full straw).

Experiment 2: Comparison of OPTEC® 5000V1 to OPTEC® 5000V2

In Experiment 2, 101 participants responded to both the OPTEC® 5000V1, the original color vision plate, and the OPTEC® 5000V2, the color vision plate covered with orange film. The instrument illuminant was the (lighting systems part # 520-49) for both administrations.

Agreement of the two versions (Tables 3 & 4) with the CAD test for diagnosis of normal or deficient color vision was essentially unchanged between the versions, Kappa $_{(n=101)}$ V1 = .564 and V2 = .563.

Table 3. Crosstabulation of CAD NCV or CVD Diagnosis by Pass/Fail on the OPTEC®5000V1

	OPTEC® 5000V1		
CAD Diagnosis			
	Fail	Pass	Total
CVD	43	8	51
NCV	14	36	50
Total	57	44	101

Table 4. Crosstabulation of CAD NCV or CVD Diagnosis by Pass/Fail on the OPTEC®5000V2

	OPTEC® 5000V2		
CAD Diagnosis			
	Fail	Pass	Total
CVD	45	6	51
NCV	16	34	50
Total	61	40	101
Kappa = .563			

Upon closer examination, Table 5 shows inconsistent individual performance (Kappa $_{(n=101)}$ =.552) between the two versions, indicating that 22 individuals (21.8%) passed one version and failed the other; Table 6 reveals that 16 of those had NCV, 3 were deutans, and 3 were both red/green and yellow/blue weak participants.

Table 5. Crosstabulation of Pass/Fail Outcome for the OPTEC®5000V1 and the OPTEC®5000V2

	OPTEC® 5000V2		
OPTEC®5000V1			
	Fail	Pass	Total
Fail	48	9	57
Pass	13	31	44
Total	61	40	101
Kappa = .552			

Table 6. Crosstabulation of the Pass/Fail Outcome for the OPTEC®5000V1 by the OPTEC®5000V2 by CAD Type Diagnosis

		OPTEC® 5000V2	
	OPTEC® 5000V1		
Normal	Fail	7	7
(n=50)	Pass	9	27
Protan	Fail	12	0
(n=12)	Pass	0	0
Deutan	Fail	24	0
(n=29)	Pass	3	2
Tritan	Fail	1	0
(n=2)	Pass	0	1
RG & YB	Fail	4	2
(n=8)	Pass	1	1

Figure 2 demonstrates that modifying the plate filter resulted in more NCV participants failing items C, D, and E, but improved performance on item F. Slightly more CVD participants failed item C on OPTEC® 5000V2 than 5000V1.

Summary of Experiments 1 and 2

Unfortunately, these two studies did not overlap such that all subjects were administered the OPTEC® 2000, 5000V1, and 5000V2, so a comparison between the 2000 and 5000V2 cannot be computed to produce an agreement statistic. More importantly, participants in both studies underwent the same diagnostic test, the CAD test; therefore, the sensitivity and specificity of each version was calculated. Although the sensitivity improved from the OPTEC® 2000 with the introduction of the OPTEC® 5000V1 in Experiment 1, it was at the expense of the test specificity, as shown in Table 7, which provides sensitivity, specificity, and Kappa for each Stereo Optical test version using the CAD as the definitive diagnostic test. Test sensitivity was good for all versions, but specificity rates were not adequate for a selection screening test with values between 58% and 72%, meaning that as many as 42% of applicants with normal color vision may fail the color vision screening test. It is important to point out that, when pilot applicants fail their initial screening test, they have the option of requesting additional testing to obtain a waiver for color vision that involves a Flight Standards District Office examiner to administer a signal light gun test, charting/map testing, and/or a medical flight test, and other testing in the airport environment, which is time-consuming and expensive for the FAA.

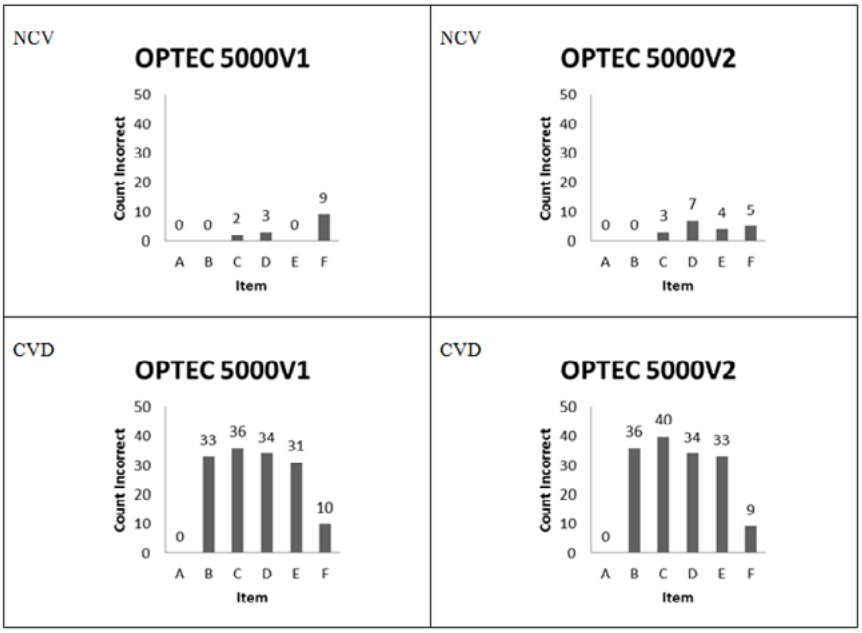

Figure 2. Count of the number of incorrect responses made by NCV (n=50) and CVD (n=51) participants in Experiment 2 comparing the OPTEC®5000V1 and the OPTEC®5000V2 by item (A-F).

Table 7. Sensitivity, specificity, and Kappa (validity) for Experiments 1 and 2 using the CAD test as the definitive diagnostic test.

	N	Sensitivity	Specificity	Kappa
Experiment 1				
OPTEC® 2000	60	87%	72%	.598
OPTEC® 5000V1	60	93%	58%	.528
Experiment 2				
OPTEC® 5000V1	101	84%	72%	.564
OPTEC® 5000V2	101	88%	68%	.563

With the specificity rates for this test, it is likely that 28 to 42% of the applicants could potentially request it, an added expense for the taxpayer. Tables 8 and 9 are crosstabulation tables of the signal light gun test with OPTEC® 5000V1 and V2 showing that agreement between the signal light gun test and the OPTEC® 5000V1 and V2 resulted in a Kappa score of .26 and .37, respectively. Based on these tables, we could predict that 54 to 65% of those that failed the OPTEC® 5000 V1 or V2 versions would pass the signal light gun test. This percentage is normally somewhat high because the color demands of color vision screening tests are typically more stringent than the signal light gun test, which employs brightness differences between red, green, and white lights, hence providing CVD examinees a redundant cue to facilitate their color naming. Conversely, only a small percentage of those who pass the OPTEC® 5000 V1 or V2 versions are likely to be *unable* to distinguish the colored lights of the SLGT, which from a safety standpoint, is a desirable screening test attribute.

Table 8. Crosstabulation of the Pass/Fail Outcome of the SLGT by the OPTEC® 5000V1 for All Subjects

	OPTEC® 5000V1		
SLGT			
	Fail	Pass	Total
Fail	35	3	38
Pass	64	58	122
Total	99	61	160
Kappa = .216			

Table 9. Crosstabulation of the Pass/Fail Outcome of the SLGT by the OPTEC®5000V2 for All Subjects

	OPTEC® 5000V2		
SLGT			
	Fail	Pass	Total
Fail	26	2	28
Pass	31	38	69
Total	57	40	97
Kappa = .367			

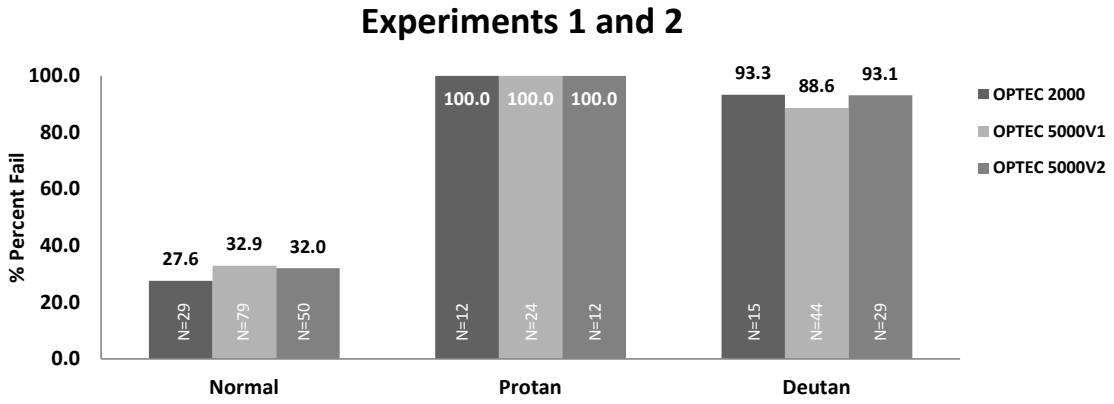

Figure 3. Percent failing by color vision type (normal, protan, deutan) for the OPTEC®2000, OPTEC®5000V1, and OPTEC®5000V2 (N indicates the total number of subjects from which the percent failing were calculated)

Figure 3 shows data for participants in Experiments 1 and 2 for the most common color vision types (normal, protan, and deutan) for the OPTEC® 2000, 5000V1, and 5000V2; the total number of subjects *from which the percent failing were calculated* are noted on each bar. Both experiments presented 5000V1; therefore, the percent failing was calculated based on the combined subject pool. One hundred percent of those with protan deficiencies failed all tests. The graph reflects the problem as the large percentage of NCV participants failing.

CONCLUSIONS

When Stereo Optical updated their 2000 model, most apparent was the esthetic design change, but the modified light source was the most crucial change, because it affected the color appearance of the pseudoisochromatic plates used for color vision screening. In their defense, few instruments were available to accurately measure LEDs' CRI because the LED technology was emerging, and scientists had not settled on an appropriate index for calculating CRI. Still, modifying the plate filter to compensate for the light source change did not improve the test as a selection/screening instrument, as we have shown in this paper.

Based on a body of published research (CIE, 1995; Davis & Ohno, 2005, 2006, 2010; Ohno, 2004, 2005; Quintero et al., 2012) on the topic of color rendering, we believe that the illuminant/light source change was responsible for the adverse effect on the test's specificity. Furthermore, we believe that exploring the color rendering of the current light source in the OPTEC® 5000 model is a good first step to verify that the test illuminant is causing the problem or, alternatively, finding a source for another illuminant with good color rendering. Making recommendations for a light source or specific indices or methods for measuring the color rendering of the LED test illuminant is beyond the scope of this paper. Regardless, in its current state, whether as originally deployed (OPTEC® 5000V1) or equipped with a modified filter (OPTEC® 5000V2), the Stereo Optical model 5000 should not be approved for aeromedical screening because of its unacceptable specificity rates and the potential for expensive additional testing that could result from NCV applicants failing.

A few last points about this and other six-item tests with regards to aeromedical screening and other safety-critical occupational screening: It is very easy for a highly-motivated examinee to memorize the correct answers to the items, especially because the first item is a demonstration plate designed for all individuals to see the numerals, and the correct answer to the last item is "blank," leaving only four trials to memorize. It is important to note that the participants in our experiments were *novices* to the tests prior to visiting the laboratory, so the sensitivity of the test reported in this paper is probably a true reflection of the test; however, CVD pilot examinees are known to "shop around" to find an aviation medical examiner who uses their preferred color vision screening test, thereby increasing their chances of passing the test. A shortcoming of six-item tests is their vulnerability to memorizing the answers. For this test, all six items can be seen at once and trials are labeled, two factors that facilitate memorization because they cannot be anonymously re-ordered. Contrast that set of circumstances to other book-based PIP tests that often involve 14 or more, un-numbered test plates that can be re-arranged or reordered to prevent memorizing responses in order of presentation.

REFERENCES

Barbur, J.L., Cole, V.A., & Plant, G.T. (1997). Chromatic discrimination in subjects with both congenital and acquired colour vision deficiencies. In Cavonius (Ed), *Colour Vision Deficiencies XIII,* 28:211-223. Kluver Academic Publishers, Dordrecht, Netherlands.

Barbur, J.L., Rodriguez-Carmona, M., & Harlow, A.J. (2006). Establishing the statistical limits of 'Normal' chromatic sensitivity. *CIE Expert Symposium, CIE Proceedings 75 Years of the Standard Colorimetric Observer*.

Code of Federal Regulations, Title 14, Part 67, Washington, DC: U.S. Government Printing Office, 2013.

Commission Internationale de l'Eclairage. "International lighting vocabulary," CIE 17.4: item 845-02-59 (1987).

Commission Internationale de l'Eclairage. "Color rendering of white LED light sources," CIE Publ. No. 177, (Central Bureau of the CIE, Vienna, Austria, 1995).

Commission Internationale de l'Eclairage. "Method of specifying and measuring color rendering properties of light sources," CIE Publ.No.13.3, (Central Bureau of the CIE, Vienna, Austrria, 1995).

Davis, W. & Ohno, Y. (2005). Toward an improved color rendering metric. In I.T. Ferguson, J.C. Carrano, T. Taguchi, & I.E. Ashdown (Eds),*Fifth International Conference on Solid State Lighting*, Proceedings SPIE 5941, 59411G.

Davis, W. & Ohno Y. (2006). Color rendering of light sources. National Institute of Standards and Technology (NIST). http://web.archive.org/web/20090505170441/http://physics.nist.gov/Divisions/Div844/facilities/vision/color.html

Davis, W. & Ohno, Y. (2009). Approaches to color rendering measurement, *Journal of Modern Optics, 56:13,* 1412–1419.

Davis, W. & Ohno, Y. (2010). Color quality scale, *Optical Engineering, 49,* 033602-033616.

Federal Aviation Administration (1992). Guide for Aviation Medical Examiners. Washington, DC: FAA Office of Aviation Medicine, pp. 80-82.

Federal Aviation Administration (2013). Guide for Aviation Medical Examiners. On line at:http://www.faa.gov/about/office_org/headquarters_offices/avs/offices/aam/ame/guide/app_process/exam_tech/item52/amd/

Mertens H.W., & Milburn, N.J. (1993). *Validity of FAA-approved color vision tests for Class II and Class III aeromedical screening*. (Report No. DOT/FAA/AM-93/17) Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Nakagawara, V.B., Montgomery, R.W., & Wood, K.J. (2009). *Evaluation of next-generation vision testers for aeromedical certification of aviation personnel*. (Report No. DOT/FAA/AM-09/13) Washington, DC: Federal Aviation Administration Office of Aerospace Medicine.

Ohno, Y. (2004). Color rendering and luminous efficacy of white LED spectra, *Proceedings of the SPIE, 5530,* 88-98.

Ohno, Y. (2005). Spectral design considerations for color rendering of white LED light sources, *Optical Engineering, 44,* 111302.

Quintero, J., Sudria, A., Hunt, C., & Carreras, J. (2012). Color rendering map: a graphical metric for assessment of illumination. *Optical Society of America, 20,* 4939-5956.

Sharpe, L.T., Stockman, A., Jagle, H., & Nathans, J. (1999). Opsin genes, cone photopigments, color vision, and color blindness. In K.R. Gegenfurtner & L.T. Sharpe (Eds.) *Color Vision: From Genes to Perception*, pp. 3-52. Cambridge, England: Cambridge University Press.

Stereo Optical Company. (Undated). Reference and instruction manual model 2000 vision tester. Chicago, IL: Stereo Optical Company, Inc.

Stereo Optical Company. (2004). Ophthalmic prescreening slide package. Chicago, IL: Stereo Optical Company, Inc.

Stereo Optical Company. (2009). Reference and instruction manual model 5000 vision tester. Chicago, IL: Stereo Optical Company, Inc.

APPENDIX A

Table A1. Response to Item A on the OPTEC® 2000 by OPTEC® 5000V1 Crosstabulation by CAD Type Diagnosis

			OPTEC 5000V1 Item A	
	Diagnosis		Incorrect	Correct
OPTEC 2000 Item A	Normal	Incorrect	0	0
		Correct	0	29
	Protan	Incorrect	1	1
		Correct	0	10
	Deutan	Incorrect	0	0
		Correct	0	15
	Tritan	Incorrect	0	0
		Correct	0	1
	RG & YB	Incorrect	1	0
		Correct	0	2

Table A2. Response to Item B on the OPTEC® 2000 by OPTEC® 5000V1 Crosstabulation by CAD Type Diagnosis

			OPTEC 5000V1 Item B	
	Diagnosis		Incorrect	Correct
OPTEC 2000 Item B	Normal	Incorrect	0	0
		Correct	1	28
	Protan	Incorrect	12	0
		Correct	0	0
	Deutan	Incorrect	10	0
		Correct	2	3
	Tritan	Incorrect	0	0
		Correct	0	1
	RG & YB	Incorrect	1	0
		Correct	0	2

Table A3. Response to Item C on the OPTEC® 2000 by OPTEC® 5000V1 Crosstabulation by CAD Type Diagnosis

			OPTEC 5000V1 Item C	
	Diagnosis		Incorrect	Correct
OPTEC 2000 Item C	Normal	Incorrect	2	0
		Correct	4	23
	Protan	Incorrect	12	0
		Correct	0	0
	Deutan	Incorrect	11	0
		Correct	2	2
	Tritan	Incorrect	0	0
		Correct	1	0
	RG & YB	Incorrect	1	0
		Correct	0	2

Table A4. Response to Item D on the OPTEC® 2000 by OPTEC® 5000V1 Crosstabulation by CAD Type Diagnosis

			OPTEC 5000V1 Item D	
	Diagnosis		Incorrect	Correct
OPTEC 2000 Item D	Normal	Incorrect	3	0
		Correct	5	21
	Protan	Incorrect	10	1
		Correct	0	1
	Deutan	Incorrect	12	0
		Correct	1	2
	Tritan	Incorrect	0	0
		Correct	0	1
	RG & YB	Incorrect	1	0
		Correct	0	2

Table A5. Response to Item E on the OPTEC® 2000 by OPTEC® 5000V1 Crosstabulation by CAD Type Diagnosis

			OPTEC 5000V1 Item E	
	Diagnosis		Incorrect	Correct
OPTEC 2000 Item E	Normal	Incorrect	1	0
		Correct	0	28
	Protan	Incorrect	11	0
		Correct	1	0
	Deutan	Incorrect	8	0
		Correct	5	2
	Tritan	Incorrect	0	0
		Correct	0	1
	RG & YB	Incorrect	1	0
		Correct	0	2

Table A6. Response to Item F on the OPTEC® 2000 by OPTEC® 5000V1 Crosstabulation by CAD Type Diagnosis

			OPTEC 5000V1 Item F	
	Diagnosis		Incorrect	Correct
OPTEC 2000 Item F	Normal	Incorrect	1	3
		Correct	3	22
	Protan	Incorrect	1	0
		Correct	4	7
	Deutan	Incorrect	2	1
		Correct	2	10
	Tritan	Incorrect	0	0
		Correct	0	1
	RG & YB	Incorrect	0	0
		Correct	0	3