Doc
FAA
AM
93
17

[illegible]-93/17

[illegible]tion Medicine
[illegible] D.C. 20591

Validity of FAA-Approved Color Vision Tests for Class II and Class III Aeromedical Screening

RECEIVED
OCT 20 1993
FAA LIBRARY

Henry W. Mertens
Nelda J. Milburn

Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

September 1993

Final Report

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation
Federal Aviation Administration

Doc
FAA
AM
93/17

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-93/17		

RECEIVED OCT 26 1993 ERAU LIBRARY

4. Title and Subtitle	5. Report Date
Validity of FAA-Approved Color Vision Tests for Class II and Class III Aeromedical Screening	September 1993
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Henry W. Mertens, Ph.D. and Nelda J. Milburn, Ed.D.	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, OK 73125	11. Contract or Grant No.
12. Sponsoring Agency name and Address	13. Type of Report and Period Covered
Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, DC 20591	14. Sponsoring Agency Code

15. Supplemental Notes

This work was performed under Task AM-D-92-HRR-107

16. Abstract

All clinical color vision tests currently used in the medical examination of pilots were studied regarding validity for prediction of performance on practical tests of ability to discriminate the aviation signal colors, red, green, and white given under both day and night conditions. Those same practical tests are given to pilots with color vision deficiency who apply for a waiver of the Class II or Class III color vision standards. Subjects with varying type and degree of color vision deficiency (n=122) and subjects with normal color vision (n=120) were classified with the anomaloscope and given both practical and clinical tests. The clinical color vision tests included the American Optical Company plates (1965 and 1940 Editions), AOC-HRR plates (2nd Edition), Ishihara plates (14-, 16-, 24-, and 38-plate tests), Dvorine plates, Richmond plates, Farnsworth Lantern, School of Aviation Medicine Color Threshold Tester, Titmus Tester, Titmus II Tester, OPTEC 2000 Tester, and Keystone Orthoscope/Telebinocular test. The criterion tests required naming the colors of actual signals produced by the Aviation Signal Light, with the same test procedures and viewing distances used in actual practical tests. The Farnsworth Lantern and several plate tests were the best clinical tests for predicting ability to identify the colors of aviation signals. Individuals with color vision deficiency identified signal colors better at night than during the day. Recommendations for improving the disposition criteria of some clinical tests, and for discontinuing several obsolete tests are discussed.

17. Key Words	18. Distribution Statement
Aviation Signal Lights Color Vision Screening Color Vision Deficiency	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	14	

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

Validity of FAA-Approved Color Vision Tests for Class II and Class III Aeromedical Screening

Introduction

The color vision requirements of the Federal Aviation Regulations (FARs), for both Class II (FAR 67.15(b)(5)) and Class III (FAR 67.17(b)(3)) aeromedical certification, are the same, i.e., "Ability to distinguish aviation signal red, aviation signal green, and white." This requirement applies to commercial (Class II) and private (Class III) pilots and is based on the international use of a color coded signal system agreed upon by members of the International Civil Aviation Organization. This signal system is used throughout the world by air traffic control (ATC) personnel in control towers to direct aircraft that have experienced radio failure or aircraft that have no radio. The signaling device is called the Aviation Signal Light. Interpretation of the code requires correct identification of the colors red, green, and white, which have different meanings depending on whether the signal light is steady or flashing, and whether the aircraft is in the air or on the ground (FAR 91.77), as shown in Table 1. An error in color perception can lead to an erroneous decision by the pilot and an increased risk of conflict with other aircraft.

Table 1
Meaning of ATC Light Signals

COLOR AND TYPE OF SIGNAL	MEANING WITH RESPECT TO AIRCRAFT ON THE SURFACE	MEANING WITH RESPECT TO AIRCRAFT IN FLIGHT
Steady green	Cleared for take-off	Cleared to land
Flashing green	Cleared to taxi	Return for landing (to be followed by steady green at proper time)
Steady red	Stop	Give way to other aircraft and continue circling
Flashing red	Taxi clear of runway in use	Airport unsafe--do not land
Flashing white	Return to starting point on runway	Not applicable
Alternating red and green	Exercise extreme caution	Exercise extreme caution

If a pilot undergoing a Class II or Class III aeromedical examination by an Aviation Medical Examiner (AME) is found to have deficient color vision, that person is automatically restricted from (1) night flight and (2) flight under color signal control, i.e. flight without a working radio into a controlled airport (day or night) where the tower would use color-coded signals from the Aviation Signal Light (ASL) to direct the aircraft.

The color deficient pilot, however, may seek a waiver of the standard. For both Class II and Class III medical certification, the waiver will typically be based on demonstration of ability to correctly identify the aviation signal colors in a practical test called the Aviation Signal Light (ASL) test. The ASL test is performed at an airport and involves identification of actual ASL signals shown from the air traffic control tower. The ASL test is first given during daylight, since reduced signal/background contrast makes it a more difficult observing condition than nighttime (Steen & Lewis, 1972; Steen, Collins, & Lewis, 1974). If the examinee passes the day ASL test, all restrictions related to color vision are removed. If the day test is failed, the examinee may request an ASL test at night. If the night test is passed, the restriction from flight at night is removed, but the restriction from flight under color signal control in daytime remains. Since the ASL test is a practical test serving as the ultimate criterion for meeting the Class II and Class III color vision standards, it is also appropriate for use as a criterion in evaluating the clinical tests used by AMEs for color vision screening for those standards.

Previous research at the Civil Aeromedical Institute (CAMI) has used the ASL test as a criterion for evaluating predictive validity of several clinical color vision tests accepted by the Federal Aviation Administration for use by AMEs (Steen, et al. 1974; Jones, Steen, & Collins, 1976), but not all accepted tests were evaluated, and several tests have been accepted since that earlier research. The present research also used the day and night ASL tests as criteria, but validated all clinical color vision tests accepted for aeromedical screening and known to be in use. Recommendations based on the present research were made for improvement of aeromedical color vision screening procedures with the clinical color vision tests, and those recommendations were incorporated in the current, 1992 edition of the FAA's *Guide for Aviation Medical Examiners.* Estimates of test reliability were assessed for the clinical color vision tests since low reliability of a predictor test may cause low validity.

Method

Subjects

The present study involved a total of 242 subjects, including 120 normal trichromats (normal color vision), and 122 individuals with red-green color vision deficiencies of varying magnitude and of both protan (red weak) and deutan (green weak) types of deficiencies. The normal trichromats had a mean age of 32 years (S.D.=9 years), and color deficients had a mean age of 30 years (S.D.=10 years). Thirty-three of the normal trichromats and 7 of the color deficients were women. Eighty-seven of the normal trichromats and 115 of color deficients were men. All subjects had at least 20/30 visual acuity in both near and distant vision, as measured by the Titmus Vision Tester or the Bausch and Lomb Orthorater vision tester. Subjects were recruited through advertisements in newsletters at Tinker Air Force Base, area colleges and universities, and through local newspapers of the Oklahoma City (OK) metropolitan area. All subjects were paid an hourly wage. During the testing sessions of this study, the subjects also participated in several tasks that were performed for two other experiments concerning evaluation of color vision requirements and selection testing for Air Traffic Control Specialists that have been reported previously by Mertens and Milburn (1992a, 1992b).

Diagnostic Color Vision Tests.

The principal diagnostic instrument was the Nagel Type I anomaloscope (Schmidt-Haensch). The anomaloscope is generally recognized as the best instrument for differentiation of normal trichromats from individuals with red-green color vision deficiencies, for differentiation of protan and deutan types among the red-green deficients, and also recommended for diagnosis and differentiation of level of deficiency (i.e. simple anomalous, extreme anomalous, and dichromats, in order of severity) by

the National Research Council-National Academy of Science (NRC-NAS) Committee on Vision (1981). Other tests were given to detect and diagnose the rare tritan or "blue-yellow type" of deficiencies that the Nagel Type I anomaloscope does not detect. A detailed description of diagnostic methods is found in Mertens and Milburn (1992a).

The anomaloscope testing procedure described by Steen, et al. (1974) was used to obtain matches for both "neutral" and "chromatic" adaptation conditions. The anomaloscope diagnostic classification procedure used was that described and recommended by the NRC-NAS Committee on Vision (1981). This procedure bases diagnoses primarily on measurements obtained under chromatic adaptation (i.e., matches obtained after observing the anomaloscope stimulus for at least 20 s). The classifications obtained are listed in ascending order of severity:

Normal. Normals (n=120) comprise the majority of normal trichromats and have a high level of color discrimination ability. The anomaloscope color matching behavior of normal trichromats varies by a relatively small amount, but it does vary. For purposes of this study, this classification also contains those normal trichromats called "deviant normal trichromats" and "weak normal trichromats." These subgroups of normal trichromats may be thought of as representing the "tails" of either the distribution for the matching range size (the "weak" normal trichromats) or the distribution for matching range midpoint (the "deviant" normal trichromats). Some of these individuals may show a very slight reduction in color discrimination ability.

Simple anomalous trichromats. The simple anomalous trichromats are the "mildest" of inherited red-green color vision deficiencies. They were separated into simple protanomalous (n=8) and simple deuteranomalous (n=23). These individuals may have mild to moderate impairment of color discrimination ability.

Extreme anomalous trichromats. These color deficients were separated into extreme protanomalous (n=18) and extreme deuteranomalous (n=25). These individuals have severe impairment. The extreme protanomalous are also distinguished by having reduced sensitivity to long wavelength (red) light.

Dichromats. Dichromats were similarly separated into protan and deutan groups called protanopes (n=18) and deuteranopes (n=30), respectively. All dichromats have severe color deficiencies and protanopes, like the extreme protanomalous, have reduced sensitivity to long wavelengths, while deuteranopes do not.

The Farnsworth Dichotomous Test for Color Blindness Panel D-15, the Farnsworth F2 plate, and the AOC-HRR test were used to detect the rare tritan deficiencies. Those tests have been described by others (NRC-NAS Committee on Vision, 1981; Pokorny, Smith, Verriest, and Pinckers, 1979). No tritans were found.

Aviation Signal Light (ASL) Test

In both daytime and nighttime ASL tests, each subject identified signal colors when positioned at distances of 1000 and 1500 feet from the location of the ASL, with three signals observed at each distance for a total of six signals in each test. The ASL signal colors (red, green, and white) were presented randomly in both day and night tests, with the restriction that each color was presented at least once during the six trials of a given test. Signals were of 5-s duration, and intervals between signals were at least 3 min. Those procedures are described in the Federal Aviation Administration (FAA) Order 8710.4 (1986). ASL signals were shown out of a third-floor, north window containing clear glass, and at an elevation of 38 feet. The daytime tests were administered between the hours of 10:00 am and 3:00 pm. Sky conditions ranged from clear to overcast during daytime ASL tests; heavy overcast was rare. The nighttime tests were administered after dark between 9:30 pm and 10:30 pm. Any error in identifying the color of one of the six ASL signals constituting a test caused failure of that test. No feedback regarding whether responses were correct or incorrect was given during the test. The ASL used in this study was the Type W-1 model manufactured by Crouse-Hinds Company (Syracuse, N.Y.).

Aeromedical Color Vision Screening Tests

All aeromedical color vision screening tests known to be in use by AMEs for the medical examination of pilots were evaluated. These tests can be grouped into three categories: 1) pseudoisochromatic plate tests, 2)

Table 2

Pre-1992 List of FAA-Accepted Color Vision Tests

ACCEPTED TESTS	FULL TEST NAME	SOURCE
Pseudoisochromatic Plate Tests		
1. AOC (18-Plate) **	American Optical Company Pseudoisochromatic Plates - 1940 Edition	Out of Print
2. AOC (15-Plate) *	American Optical Company Pseudoisochromatic Plates - 1965 Edition	Out of Print
3. AOC-HRR	American Optical - Hardy, Rand, Rittler Pseudoisochromatic Plates - 2nd Ed.	Out of Print
4. Dvorine (2nd Ed.)	Dvorine Pseudo-Isochromatic Plates - 2nd Ed.	Psychological Corp. 555 Academic Court San Antonio, TX 78204-0954
5. Ishihara (14-Plate) *	Ishihara's Tests for Color-Blindness Concise Edition	Graham-Field Surgical Co., Inc. New Hyde Park, N.Y. 11040
6. Ishihara (16-Plate)	16 Plates Edition (Out of Print)	
7. Ishihara (24-Plate)	24 Plates Edition	
8. Ishihara (38-Plate)	38 Plates Edition	
9. Richmond *	Richmond Pseudo-Isochromatic Plates for Testing Color Perception	Richmond Products, Inc. 1021 S. Rogers Circle Boca Raton, FL 33487
Lantern Tests		
10. Farnsworth Lantern	Farnsworth Lantern	Macbeth Corp. P.O. Box 230 Newburgh, NY 12551
11. SAMCTT **	School of Aviation Medicine Color Threshold Tester	Out of Production (Formerly from Macbeth Corp.)
12. Edridge-Green Lantern **	Edridge-Green Lantern	Rayner Optical Co., Ltd. 17 Lorna Road, Hove East Sussex, BN3 3EP United Kingdom
Multifunction Vision Testers		
13. Titmus	Titmus Vision Tester	Titmus Optical, Inc. 1015 Commerce St. Petersburg, VA 23803
14. Titmus II *	Titmus II Vision Tester	
15. OPTEC 2000 *	OPTEC 2000 Vision Tester, Aeromedical Model	Stereo Optical Co. 3539 Kenton Ave. Chicago, IL 60641
16. Keystone Orthoscope/ Telebinocular	Keystone Telebinocular Test Unit, Aeromedical	Keystone Co. 4673 Air Center Circle Reno, NV 89502

* Not listed in pre-1992 editions of the AME Guide
** Not listed in 1992 edition of the AME Guide

lantern tests, and 3) multifunction vision testers. The tests included those listed in previous editions of the FAA *Guide for AMEs*, and the several new tests that were accepted by the Office of Aviation Medicine but were not listed in the *Guide for AMEs* prior to the current, 1992 edition. The full names of all tests and their sources are provided in Table 2. The disposition criteria for all tests are given in Table 3; both pre-1992 and current disposition criteria listed in the 1992 edition of the *Guide for AMEs* are given.

Pseudoisochromatic Plate Tests. The 14-, 16-, and 24-plate Ishihara tests all involve subsets of plates from the 38-plate Ishihara test. Only the 14-plate and 38-plate Ishihara tests were administered. Scores for the 16- and 24-plate tests were computed based on responses to appropriate subsets of plates of the 38-plate test. It is assumed that any differences between the plates of this simulation and the actual plates of those tests are minor and would be similar to different printings of the same plate. The Dvorine, Richmond,

Table 3
Pre-1992 and Current FAA Recommended Disposition Criteria for Class II and Class III Aeromedical Color Vision Screening

ACCEPTED TESTS	PRE-1992 FAA RECOMMENDED FAILURE CRITERIA	CURRENT FAILURE CRITERIA
Pseudoisochromatic Plate Tests		
1. AOC (18-Plate)	13 or more errors, plates1-18	dropped from list
2. AOC (15-Plate) *	5 or more errors, plates 1-15	7 or more errors
3. AOC-HRR	any error on plates 12-14	any error on plates 7-11
4. Dvorine (2nd Ed.)	12 or more errors, plates 1-15	7 or more errors
Ishihara's Tests for Color-Blindness		
5. Ishihara (14-Plate) *	2 or more errors, plates 1-11	6 or more errors
6. Ishihara (16-Plate)	4 or more errors, plates 1-8	4 or more errors
7. Ishihara (24-Plate)	7 or more errors, plates 1-15	7 or more errors
8. Ishihara (38-Plate)	9 or more errors, plates 1-21	9 or more errors
9. Richmond *	5 or more errors, plates 1-15	5 or more errors
Lantern Tests		
10. Farnsworth Lantern	average of more than 1 error in series 2 and 3	average of more than 1 error in series 2 and 3
11. SAMCTT	score of 49 or less	dropped from list
12. Edridge-Green Lantern	any error, test A	dropped from list
Multifunction Vision Testers		
13. Titmus	any error	any error
14. Titmus II *	any error	any error
15. OPTEC 2000 *	any error	any error
16. Keystone Orthoscope/ Telebinocular	any error	any error

* Not listed in Pre-1992 editions of the AME Guide

and AOC-18 plate, -15 plate and -HRR were also given. A Macbeth Easel Lamp was used to illuminate all plate tests, with all other lighting extinguished during testing. All screening tests were administered according to manufacturer's directions. Among the above tests, the Ishihara 14-Plate, Richmond, and AOC 15-Plate tests were not listed prior to 1992.

Lantern Tests. The Farnsworth Lantern test was developed by the U.S. Navy and is currently used for aeromedical screening by both the Navy and the U.S. Air Force. The Farnsworth Lantern test was given in a normally lighted room, according to the manufacturer's directions, with the subject positioned 8 ft in front of the lantern. The U.S. Air Force's School of Aviation Medicine Color Threshold Tester (SAM-CTT) was given in a dark room according to the manufacturer's instructions. The Edridge-Green Lantern test was the only FAA-accepted test that was not given. That test is believed nonexistent among AMEs. Furthermore, we could not find any evidence that there is an accepted standard procedure for administration of that test, which has complex controls for color, size, and brightness of test stimuli.

Multifunction Vision Testers. Several multifunction vision testers include color vision tests that are accepted by the FAA; all involve photographic reproductions, either prints or positive transparencies, of pseudoisochromatic plates. The Titmus, Titmus II, and OPTEC 2000 vision testers all involve positive color transparencies, which are reproductions of Ishihara test plates. The Keystone test involves photographs of six plates that were developed specifically for that test. The Keystone test was illuminated using only that instrument's lamp; room lights were turned off. Among the tests involving vision testers, the Titmus II and OPTEC 2000 tests were not listed in the FAA *Guide for AMEs* prior to the current, 1992 edition.

Procedure

The diagnostic tests and clinical color vision tests were administered at four testing stations, each supervised by a trained laboratory technician. All anomaloscope testing was performed by the first author. The tests at each station took approximately 45 min to administer. The testing of each subject was performed in two, 2-hour sessions separated by a 1-hour lunch break or given on successive days. A separate session for the night ASL test lasted approximately 1 hour. Within each daytime session, the two 45-min testing periods were separated by a 15-min break. Since the OPTEC 2000, Titmus, and Titmus II involved reproductions of the same set of six Ishihara plates, each of those tests was placed at a different testing station to separate them in time and by administrations of several other tests. Four subjects were usually tested simultaneously but separately, one at each of the four test stations. Upon reporting for the experiment, each subject was assigned a prearranged random order for testing at the four stations. Six subjects did not return for the nighttime ASL test. Approximately two-thirds of subjects had the night ASL test after the day ASL test; prior research by Steen and Lewis (1972) found that order of day and night ASL testing had no effect on performance.

Results and Discussion

Relation of Performance on ASL Tests to Type and Degree of Color Vision Deficiency

The number of subjects passing and failing day and night ASL tests is presented in Table 4 as a function of type and degree of color vision deficiency. Failure was rare among normal trichromats; only one normal failed a day ASL test and no normals failed the night ASL test. In contrast, many simple and extreme anomalous trichromats, and dichromats made errors in both day and night ASL tests. In agreement with earlier findings of Steen, et al. (1974), the proportion of deutans that passed the day ASL test was larger (36%) than among the protans (11%). That tendency was also apparent in the night ASL test, 65% and 40% passed among deutans and protans. The finding that more subjects with color vision deficiency passed the ASL test at night than in daytime, 56% versus 27%, respectively, confirms earlier findings that ASL signal colors are easier to identify at night for many color deficient observers. The above findings are in agreement with the results of previous CAMI research (Steen & Lewis, 1972; Steen, et al., 1974) in showing that a substantial proportion of color deficient individuals cannot reliably

Table 4
Number of Subjects Passing or Failing the ASL Test in Daytime and Nighttime as a Function of Anomaloscope Classification

		NORMAL	ANOMALOUS TRICHROMATS				DICHROMATS	
			Simple		Extreme			
			Prot	Deut	Prot	Deut	Prot	Deut
Daytime	Pass	119	3	15	2	6	0	7
	Fail	1	5	8	16	19	18	23
Nighttime	Pass	117	4	17	6	18	7	15
	Fail	0	3	5	12	7	10	15

identify aviation signal colors. They also confirm that a number of color deficient individuals retain sufficient color vision ability to differentiate aviation signal colors.

Relation of ASL Test Performance to Performance of Clinical Color Vision Tests

The validities of color vision tests were evaluated regarding prediction of pass/fail performance on the daytime and nighttime ASL tasks. Cohen's Kappa (k), an index of agreement, was used to assess the validity of the color vision screening tests, as recommended by the NAS-NRC Committee on Vision (1981). The index can be interpreted as the percentage agreement between test and criterion variable, with correction for chance. In addition to validities, each table lists the miss and false alarm rates for each test. *Miss rate* is the probability of passing the clinical test given that the criterion ASL test will be failed. The *false alarm rate* is the probability of failing the clinical test given that the criterion ASL test will be passed. *Miss* and *false alarm rates* are sometimes alternatively referred to as false negative rate and false positive rate, respectively.

The validities for clinical tests using pre-1992 disposition criteria (shown in Table 3) for prediction of day and night ASL test performance are shown in Table 5 along with miss and false alarm rates. The validity of screening tests was also compared with the validity of an anomaloscope classification of "Normal" versus deficient color vision as a predictor of performance. A diagnosis of normal trichromat, as defined in Mertens and Milburn (1992a), was required for "passing" the anomaloscope test in the present analysis. Passing the anomaloscope was slightly less effective as a predictor of day ASL test performance (k=.72) than the better clinical tests which had validities as high as k=.79 in the case of the Farnsworth Lantern test.

Several tests had substantially lower validities with the pre-1992 disposition criteria, and most importantly, unacceptable high miss rates. Those tests were the AOC-HRR, the Dvorine, and the AOC (18-Plates) tests. The unacceptable miss rates ranged from 18.2 to 54.4%. Several tests had acceptable miss rates, but had false alarm rates above 20%. In many practical situations, including the aviation situation, a miss is considered as more important than a false alarm. Jones, et al. (1975) discussed this issue as follows:

> "For example, if a miss occurs in aviation, a pilot may be allowed to fly under circumstances for which he [*sic*] may not be qualified; however, if a false alarm occurs, he [*sic*] can request a 'second chance,' i.e., a chance to demonstrate his [*sic*] ability to discriminate color in a practical situation. Therefore, in the comparison of tests, an 'acceptable' miss rate, or the probability of a miss, should be selected and then the false alarm rates for the tests at that miss rate should be compared. The test with the lower false alarm rate for a given miss rate is the better predictor of performance..."

The better predictors of day ASL performance in the present research, such as the Farnsworth Lantern, had a miss rate of 5% or less and a false alarm rate of 20% or less. In evaluating alternative disposition criteria for tests, we emphasized achieving a miss rate of 5% or less, and secondarily, minimizing the false alarm rate below 20%. We did not attempt to find alternative

Table 5

Validity of Color Vision Selection Tests for Prediction of Performance on Day and Night ASL Tests using Pre-1992 Disposition Criteria

	DAYTIME ASL			NIGHTTIME ASL		
	ERROR RATE (%)			ERROR RATE (%)		
Current Tests	Miss	False Alarm	Validity (Kappa)	Miss	False Alarm	Validity (Kappa)
Anomaloscope	1.1	21.7	.72	0.0	36.4	.43
AOC (18-Plate)	40.0	7.9	.55	36.5	17.4	.42
AOC (15-Plate)	0.0	25.7	.68	0.0	40.2	.40
AOC-HRR	28.9	7.9	.65	30.8	21.2	.41
Dvorine	25.6	9.2	.40	25.0	21.7	.45
Ishihara (14-Plate)	0.0	20.4	.74	0.0	36.4	.43
Ishihara (16-Plate)	0.0	19.1	.76	0.0	35.3	.45
Ishihara (24-Plate)	0.0	19.1	.76	0.0	35.3	.45
Ishihara (38-Plate)	0.0	18.4	.77	0.0	34.8	.45
Richmond	0.0	23.0	.71	0.0	38.6	.41
Farnsworth Lantern	5.6	13.2	.79	7.7	29.9	.46
SAMCTT	10.0	11.9	.77	5.8	25.5	.53
Titmus Tester	0.0	61.2	.32	0.0	68.5	.17
Titmus II Tester	0.0	56.6	.36	0.0	65.8	.19
OPTEC 2000	0.0	41.4	.51	0.0	52.2	.29
Keystone	0.0	24.3	.70	0.0	39.7	.40

disposition criteria for the vision tester tests that utilize less stable photographic media for test plates. In the absence of spectrophotometric control of the test materials, which has been recommended by the NAS-NRC Committee on Vision (1981) for tests using photographic materials, we recommend that their disposition criteria continue to be strict, i.e. fail with any error.

Improved Disposition Criteria for Prediction of Day ASL Test Performance

Disposition criteria were varied for clinical color vision tests to obtain optimal decision criteria for prediction of day ASL performance, with a few exceptions. Those tests not included in this analysis were the four above-mentioned tests made with photographic materials, and three obsolete tests discussed below: the AOC (18-Plate), the SAM-CTT, and the Edridge-Green lantern. The objective of this analysis was, most importantly, to achieve a miss rate of 5% or less, as discussed above. Advantageous changes in disposition criteria were found for the AOC (15-Plates), AOC-HRR, the Ishihara (14-Plates), Dvorine, and Richmond tests. The disposition criteria that we currently recommend, including the new, revised disposition criteria for several tests, are shown in the right column of Table 3. The validities obtained with the currently recommended disposition criteria are shown in Table 6. In prediction of day ASL test performance, validities of k=.75 or greater were obtained for all clinical color vision tests except the vision tester tests. Miss rates are acceptable for all tests, and are 5% or less. False alarm rates below 20% were obtained for all tests except the vision tester tests.

Jones, et al. (1975) previously performed a similar analysis for the AOC (15-Plate) and Dvorine tests to determine optimal disposition criteria for prediction of day ASL test performance. Their results corroborate our findings regarding the two tests. The optimal disposition criterion they recommended for both tests was identical to our recommendation: "fail with 7 or more errors."

Jones, et al. also found error rates for the AOC (15-Plate), Dvorine, and Farnsworth lantern that are very similar to our findings.

Disposition Criteria for Prediction of Night ASL Test Performance

Validities for prediction of night ASL test performance by clinical color vision tests scored with pre-1992 disposition criteria are shown in Table 5, and validities obtained with the recommended decision criteria are given in Table 6. The validity of clinical tests for prediction of night ASL performance were generally much lower than for day ASL performance due to higher false alarm rates.

The recommended disposition criteria were designed to optimize the prediction of daytime performance and did not always raise validity of night ASL test performance. The observed changes in validity for prediction of night ASL test were generally small, and experimenting with different disposition criteria selected specifically to increase prediction of night performance was not productive. It was not possible to increase validity for predicting night ASL test performance above k=.54 with any clinical color vision test. Any gains in validity for prediction of night ASL test performance by using a different, optimal criterion were too small to be worthwhile, in our judgment.

A Recent Change in ASL Test Procedure

Following completion of data collection for this study, we received FAA Order 8700.1 (FAA, 1992) which is a replacement for FAA Order 8710.4 (FAA, 1986). Order 8700.1 effectively doubles the number of signals shown in the ASL test. The older procedure of Order 8710.4 required random presentation of all 3 signal colors within 6 trials, 3 at 1,000 ft and 3 at 1,500 ft, such that each color must be presented at least once during the test. Order 8700.1 requires random presentation of all 3 colors at both distances, such that each color will be presented at least twice during the test, once at each of the two distances. All other aspects of the test procedure are unchanged. The increase in

Table 6

Validity of Color Vision Selection Tests for Prediction of Performance on Day and Night ASL Tests Using Current Decision Criteria

	DAYTIME ASL			NIGHTTIME ASL			ESTIMATED RELIABILITY OF TEST
	ERROR RATE (%)			ERROR RATE (%)			
Accepted Tests	Miss	False Alarm	Validity (Kappa)	Miss	False Alarm	Validity (Kappa)	(Alpha)
AOC (15-Plate)*	2.2	18.4	.75	0.0	33.7	.46	.93
AOC-HRR*	1.1	16.4	.78	0.0	32.6	.48	.78
Dvorine*	1.1	18.4	.76	0.0	34.2	.46	.96
Ishihara (14-Plate)	2.2	17.8	.76	0.0	33.2	.47	.97
Ishihara (16-Plate)	0.0	19.1	.76	0.0	35.3	.45	.97
Ishihara (24-Plate)	0.0	19.1	.76	0.0	35.3	.45	.98
Ishihara (38-Plate)	0.0	18.4	.77	0.0	34.8	.45	.98
Richmond*	2.2	16.4	.77	0.0	32.1	.18	.93
Farnsworth Lantern	5.6	13.2	.79	7.7	29.9	.46	.96
Titmus Tester	0.0	61.2	.32	0.0	68.5	.17	.86
Titmus II Tester	0.0	56.6	.36	0.0	65.8	.19	.90
OPTEC 2000	0.0	41.4	.51	0.0	52.2	.29	.92
Keystone	0.0	24.3	.70	0.0	39.7	.40	.90

*Revised Disposition Criterion

the number of trials in the ASL test would be expected to increase its reliability, and that increase in reliability of the criterion test would be expected to increase the validity of predictor tests (Gulliksen, 1987). Therefore, while the validities of the better clinical tests found in the present study are good, k=.75-.79, somewhat higher validities would be expected with the new ASL test procedure.

Estimated Reliability of Clinical Color Vision Tests.

An ideal evaluation of reliability of a clinical color vision test would involve test-retest reliability. In the present case, time and cost limitations did not permit each test to be given twice. Therefore, the reliability of each test was estimated using Cohen's (1960) alpha. Alpha provides an estimate of reliability based on the average intercorrelation among the items of a test and the number of items (Nunnally, 1978). Alpha should also be considered an estimate of the upper limit of the reliability. The values of alpha are given in Table 6 for all clinical color vision tests. All tests have respectable reliability; most tests have alpha values above 0.90, which suggests very good reliability. There is general agreement between estimates of reliabilities obtained in our research and test-retest reliabilities obtained by the NRC-NAS Committee on Vision (1981) for several clinical color vision tests. The Committee on Vision analyzed the data of others to obtain kappa as the index of reliability use of the tests in screening for normal color vision. Test-retest reliability for the AOC (15-Plates) was k=.96, for the SAM-CTT, k=.94, and for the Farnsworth Lantern, k=.98. Two recent studies compared performance of the Ishihara (24-Plates) (Johnson, 1992) and Dvorine tests (Milburn & Mertens, in press) with both the Macbeth Easel Lamp, the preferred illuminant according to the NRC-NAS Committee on Vision (1981), and another different illuminant proposed as a substitute for the Macbeth Easel Lamp. Since those studies found performance essentially the same under the different illuminants, the kappas for agreement of performance under the two illuminants can be considered to estimate test-retest reliability. Milburn and Mertens (in press) found k=.97 for the Dvorine, and the data of Johnson (1992) yield k=1.00 for the Ishihara (24-Plates). The similarity of the above-mentioned test-retest reliabilities with alphas obtained in the present study lends confidence to the use of alpha for estimating reliability of color vision tests.

Obsolete Tests

We recommend that three tests be dropped from the list of color vision tests accepted for aeromedical screening. First, the Edridge-Green lantern test should be dropped for the same reasons that we did not evaluate it in this study; it is no longer used, to our knowledge, and there is no standardized test procedure. Second, during the period of the present research, we learned that the U.S. Air Force, the only user of the SAM-CTT, has discontinued use of that test. Thus, the SAM-CTT should also be dropped because there are no users. Third, the AOC (18-plate) test should be dropped because those tests are now over 50 years old and there are very few copies that remain in use. Our own copy is also so worn and faded that we question generalization of its performance to other copies.

Conclusions

1. The findings of this study confirm the importance of a color vision requirement for pilots. Many pilots with color vision deficiency cannot reliably identify the aviation signals used by ATC in the event of radio failure.
2. These findings also confirm that a number of color deficient individuals retain sufficient ability to differentiate aviation signal colors. These findings, therefore, also support the need to offer a practical ASL test for secondary screening of pilot applicants. The availability of the ASL test prevents rejection of those applicants with color vision deficiency who can correctly identify aviation signals.
3. The better FAA-accepted clinical color vision tests had validities of k=.75-.79 for prediction of daytime identification of ASL signal colors.
4. The estimated reliabilities of the clinical color vision tests, in general, were very good. This finding suggests that the FAA-accepted clinical color vision tests will provide consistent performance in aeromedical screening.

References

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

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

Federal Aviation Administration. (1986). *Certification: Pilots and Flight Instructors.* Order No. 8710.4. (Chapter 3, Paragraph 61) Washington, DC: FAA Flight Standards Service.

Federal Aviation Administration. (1992). *General Aviation Operations Inspector's Handbook.* Order No. 8700.1, Change 8.(Volume 2, Chapter 27, Section 7). Washington, DC: FAA Flight Standards Service.

Gulliksen, H. (1987). *Theory of mental tests.* Hillsdale, NJ: Lawrence Earlbaum. (pp. 75-107)

Johnson, D. D. (1992). The True Daylight Illuminator (TDI): A less expensive source of illumination for color vision screening. *Journal of American Optometry Association*, 63, 491-495.

Jones, K. N., Steen, J.A., & Collins, W. E. (1975). Predictive validities of several clinical color vision tests for aviation signal light gun performance. *Aviation, Space, and Environmental Medicine*, 46, 660-667.

Mertens, H. W., & Milburn, N. J. (1992a). Performance of color-dependent tasks of Air Traffic Control Specialists as a function of type and degree of color vision deficiency.Report No. DOT/FAA/AM-92/28. Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Mertens, H. W., & Milburn, N. J. (1992b). Validity of clinical color vision tests for Air Traffic Control Specialists. Report No. DOT/FAA/AM-92/29. Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Milburn, N. J., & Mertens, H. W. (1993). Validation of an inexpensive test illuminant for aeromedical color vision screening. Report No. DOT/FAA/AM-93/16. Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

National Research Council-National Academy of Sciences, Committee on Vision. (1981). (pp. 9-11, 18-19, 40-78, 82) *Procedures for testing color vision: Report of working group 41.* Washington, DC: National Academy Press.

Nunnaly, J. C. (1978). *Psychometric theory.* New York: McGraw Hill.

Pokorny J., Smith V. C., Verriest G., & Pinckers, A. J. L. G. (1979). *Congenital and acquired color vision defects.* (pp. 83-106) New York: Grune and Stratton.

Steen, J. A., Collins, W. E., & Lewis M. F. (1974). Utility of several clinical tests of color defective vision in predicting daytime and nighttime performance with the aviation signal light gun. *Aerospace Medicine*, 45, 467-472.

Steen, J. A., & Lewis, M. F. (1972). Color defective vision and day and night recognition of aviation color signal light flashes. *Aerospace Medicine*, 43, 34-36.