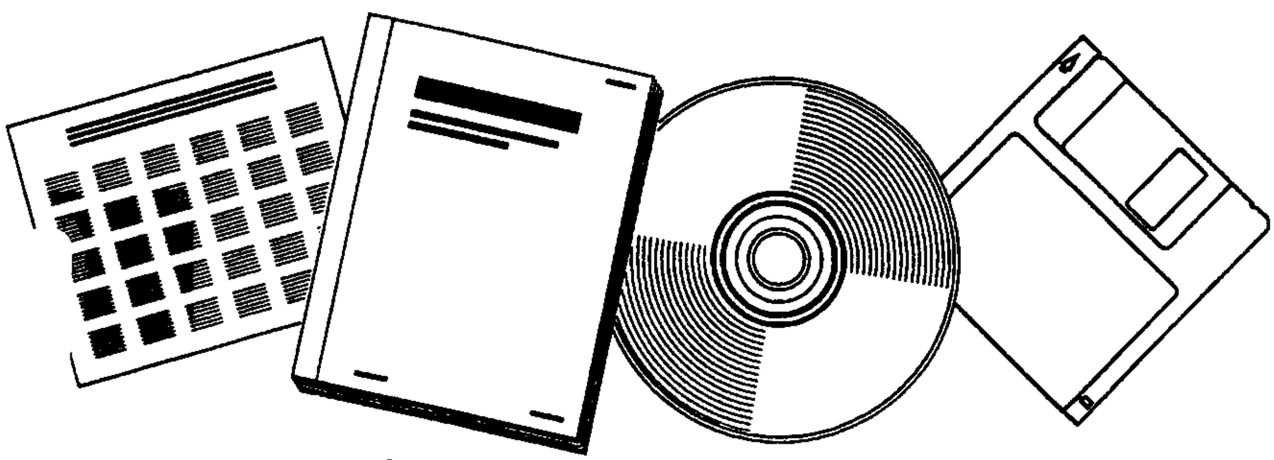
ADA132649

COLOR PERCEPTION AND ATC JOB PERFORMANCE

FEDERAL AVIATION ADMINISTRATION,
WASHINGTON, DC. OFFICE OF AVIATION
MEDICINE

JUL 1983



'ARTMENT OF COMMERCE
Technical Information Service

AD A132649

FAA-AM-83-11

ADA 132 649

COLOR PERCEPTION AND ATC JOB PERFORMANCE

Evan W. Pickrel and John J. Convey
Office of Aviation Medicine
Federal Aviation Administration

JULY 1983

Prepared for
U.S. DEPARTMENT OF TRANSPORTATION
Federal Aviation Administration
Office of Aviation Medicine
Washington, D.C. 20591

Technical Report Documentation Page

1. Report No.: FAA-AM-83-11

2. Government Accession No.: AD-A132 649

3. Recipient's Catalog No.:

4. Title and Subtitle: Color Perception and ATC Job Performance

5. Report Date:

6. Performing Organization Code:

7. Author(s): Evan W. Pickrel and John J. Convey

8. Performing Organization Report No.:

9. Performing Organization Name and Address:
Office of Aviation Medicine
Federal Aviation Administration
800 Independence Ave., SW
Washington, D.C. 20591

10. Work Unit No. (TRAIS):

11. Contract or Grant No.:

12. Sponsoring Agency Name and Address:
Office of Aviation Medicine
Federal Aviation Administration
800 Independence Ave., SW
Washington, D.C. 20591

13. Type of Report and Period Covered:

14. Sponsoring Agency Code:

15. Supplementary Notes:

Air Traffic Control Specialist

pseudoisochromatic plate

16. Abstract

Current OPM policy and guidance requires demonstrated job-relatedness and reasonable accommodation in the application of physical qualifications. The OPM has accomplished an analysis of the ATCS series and recommended development of functional color vision tests to reflect as closely as possible the functional color vision requirements of the specialty. If the PIP test is retained for prescreening to identify applicants for whom followup functional performance testing or reasonable accommodation is necessary, its use also must be standardized.

This research is directed toward accomplishment of those recommendations. A standard PIP test was validated against performance of ATCS tasks, and it demonstrated job relatedness and reasonable accommodation for application of physical qualification standards. A functional color vision test was created, but further development and validation would be needed before its operational use, and procurement would be very costly as compared to the standard PIP tests that are readily available to medical examiners.

17. Key Words: Air Traffic Controllers, color vision standards, vision, performance tests, simulation

18. Distribution Statement: Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report): Unclassified

20. Security Classif. (of this page): Unclassified

21. No. of Pages: 25

22. Price:

Form DOT F 1700.7 (8-72)

ACKNOWLEDGEMENTS

The authors thank Drs. Andrew F. Horne and Joseph A. Tucker for assistance in recruiting subjects for the study and developing the manuscript, and Nurses Irma Hart and Donna L. Hansen for administration of the tests.

COLOR PERCEPTION AND ATC JOB PERFORMANCE

Office of Personnel Management (OPM) qualification standards require that all applicants for air traffic control specialist (ATCS) positions "demonstrate normal color vision"(1). This standard has been interpreted to mean that an applicant must be able to pass any one of six different pseudoisochromatic plate (PIP) tests for color vision (2). The present practice is for aviation medical examiners (AME) to report the test used as well as the "pass" or "fail" results on the Second Class Airmen form used for the ATCS physical examination. If the vision standards for the Second Class requirement are mistakenly used, some candidates may not satisfy the standard for "normal color vision" (a technical term as defined in the literature on vision research and testing; 7 and 8). The standard for the Second Class certificate requires fewer number of correct responses than does the standard for normal color vision. At present, the testing conditions and materials used for the screening of normal color vision are lacking in standardization. Thus the scores used for pass-fail determination cannot be verified as being the same for all examinees. Standardization of the assessment of normal color vision would greatly improve the reliability of these screening measures and the validity of the decisions based on them.

Current OPM policy requires demonstrated job-relatedness and reasonable accommodation in the application of physical qualifications (3). The OPM has accomplished an analysis (4) of the ATCS series and recommended development of functional color vision tests "to reflect as closely as possible the functional color vision requirements of the specialty. If the PIP test is retained for prescreening to identify applicants for whom followup functional performance testing or reasonable accommodation is necessary, its use also must be standardized." This research is directed toward accomplishment of those recommendations. A standard PIP test was validated against performance of ATCS tasks, and it demonstrated job relatedness and reasonable accommodation for application of physical qualification standards. A functional color vision test was created, but further development and validation would be needed before its operational use, and procurement would be very costly as compared to the standard PIP tests that are readily available to medical examiners.

BACKGROUND

Color Vision Capabilities

Individual differences in the ability to perceive colors are well-known. As stated by one authority (5), "Normal persons can make visual distinctions of three types: light from dark, yellow from blue, red from green; light-dark being the most primitive type of distinction, and red-green the last acquired. Some otherwise normal persons fail to develop in their organs of sight more than a vestige of the mechanism for red-green discrimination. They are called red-green blind, or partially color-blind. A few persons have only the ability to make light-dark discrimination; they are called totally color-blind." The hereditary nature of color vision disorders was recognized at the end of the eighteenth century. Red-green color vision defects have x-chromosome-linked recessive inheritance, and are sex-linked traits. The male who inherits this defect from his mother will always have the defect, but the female must inherit it from both parents to show it. Persons with these defects comprise about ten percent of the U.S. male population and about one percent of the U.S. female population. Thus application of a color vision job standard would be discriminatory against men if it were not related to job performance, as defined in the Uniform Guidelines (6), but these defects can interfere with a person's ability to perform certain tasks satisfactorily.

Color Vision and Job Performance

There are many occupations in which defective color vision is either undesirable or unacceptable. In particular, persons who are color blind should be excluded from activities in which their confusion of color can endanger public safety. For example, In the nineteenth century there were major disasters with loss of life in the shipping and railroad industries. These tragedies often were attributed to the failure of engineers to recognize the color of a signal. As a result, people with red-green defects were, and still are, excluded from positions as pilots or engineers in commercial air, sea, and rail transport and similar duties in the armed forces. Many tests have been developed for quick, inexpensive and efficient screening of job applicants to identify those with deficient color perception.

Measuring Color Vision Capabilities

An excellent, detailed report on measurement of color vision capabilities is available elsewhere (7). Color vision tests range from those designed to make a quick identification to those for detailed diagnosis of persons with color defects. Screening tests are those designed for quick and easy identification of such persons. One of the earliest (1684) methods used to test color vision was to compare an individual's color naming of everyday objects with that of a normal person. In 1837 an advance in testing required a person to choose, from a wide range of colored samples, those that matched or most closely resembled a selected test sample. The task was performed by inspection and without color naming. Variations in the materials used among these test versions included skeins of wool, small beads or pellets, and small square pieces of colored cardboard.

Pseudoisochromatic plates were first introduced in 1873, and their success depends on the ability of color-defective persons to discriminate between certain colors. A symbol composed of colored spots is set in a background of differently colored spots, with colors chosen so that the symbol is not seen by the color-defective person. There are many modern variants of this kind of test.

The lantern test was introduced in 1903 as simulation of a working condition. It does not specifically screen for color defects, although it is expected that color-defective persons will not perform as well as those with normal color vision. An arrangements type of test was developed in 1934 that requires persons to grade and then match a series of nitrocellulose lacquer discs that vary in saturation and hue. Multiple variations of this test are available.

Diagnostic Instruments: Anomaloscopes are optical instruments that were introduced in 1881. They require a person to manipulate stimulus control knobs to match two color fields in color and brightness. They are the only clinical method for precise diagnosis of presumed genetic entities, but are most difficult to use and extensive training of examiners is necessary.

Screening Tests: For tasks from which people with major color vision defects must be excluded, a committee on vision from the National Academy of Sciences recommended the use of any of the validated screening plate tests (7). Validation here referred to demonstrated relation to performance on an anomaloscope, as a measure of presumed genetic deficiencies. Certain advantages in using the pseudoisochromatic plate tests are: (1) they are rapidly and easily administered by inexperienced personnel; (2) they are readily available and relatively inexpensive; and (3) they can be used for the general population, without accord to mental aptitude or age. They should be used primarily as screening tests to divide people into normal and color-defective populations; their diagnostic value is limited. Examples of tests include the American Optical, Dvorine, Ishihara, and other series of pseudoischromatic plates. These tests have been shown to detect about 96 percent of the cases confirmed by anomaloscope, thus validating their use in screening for presumed genetic entities. It is also important that the test(s) elected for use in screening for a particular job be validated against success on that particular job, i.e. make a determination that color deficiences as indicated by scores on the selected test(s) are related to the ability to perform job tasks.

Validation Requirements

Tests are a most visible part of the selection process, and generally are the best hope for assuring fairness and objectivity in the treatment of all applicants. The aim of testing is to identify those who are best prepared by aptitude and training to perform satisfactorily in a given role. Vision tests were initially chosen for screening because of their quite obvious content relation to successful performance in certain jobs; color vision shows this direct content relation to some air traffic controller tasks. Many kinds of tests have been and continue to be in operational use without statistical evidence to support their application. However, passage of the 1964 Civil Rights Act resulted in a renewal of interest in selection processes and especially a closer examination of those applied in high visibility career fields. The Uniform Guidelines on Employee Selection Procedures(6) has been developed and subsequently adopted by the OPM, the

Equal Employment Opportunity Commission, and other agencies having some jurisdiction over FAA hiring practice. It requires that "employer policies or practices which have an adverse impact on employee opportunities of any race, sex, or ethnic groups are illegal unless justified by business necessity".

The Uniform Guidelines provide requirements for content, construct, and criterion-related validation studies. The validation procedures for criterion-related studies must involve a statistical study; expert or professional opinion is not acceptable in lieu of a proper statistical study for this type of validity evidence. Practical significance and high utility as compared to a low amount of adverse impact is the primary requirement. A relation between performance on the job selection procedure and performance on the criterion that is statistically significant at the .05 level is considered to be an absolute minimum standard.

VALIDATION OF A PIP TEST FOR ATCS TASKS

Selecting a PIP Test

The color vision test used in this study was the Pseudoisochromatic Plates Test from the American Optical Corporation. This plate test is listed in the FAA's Guide for Aviation Medical Examiners (2) and commonly used for ATCS screening. The American Optical Corporation (AOC) marketed a 1940 edition of their PIP using 18 plates that is commonly used by designated AMEs to screen ATCS applicants. That version was superseded by a 1965 edition using 15 plates that is available at a cost of approximately $62 and described in detail elsewhere (7). Twelve plates are common to both editions. All but three plates (numbers 6, 10, 15) in the newer version were used in the 1940 edition. Six of the 18 plates in the older version (numbers 2, 3, 8, 9, 10, 13) were not included in the 1965 edition. Instructions are common to the two editions: a person is asked to "please read the numbers" and allowed about two seconds for response to each plate. When a person hesitates, the instructions are repeated once. The 1965 edition (8) was administered to all persons included in this study, and their total test scores were compared to the standards for Second Class Airmen and normal color vision performance requirements listed in the FAA Guide (2).

Selecting A Sample of ATC Tasks

The Air Traffic Control Specialist GS-2152 career field includes the options of Air Route Traffic Control Center (ARTCC) enroute controller, Airport Tower Controllers (visual and radar), and Flight Service Station (FFS) specialists. All persons listed on the register for this career field are expected to be able to perform in all options, except for some applicants restricted to the FSS option, and all options involve tasks that require normal color vision. The number and variety of ATCS tasks accomplished in the various air traffic control facilities are great. In 1981 the OPM accomplished an examination of ATCS job requirements for color vision, asking nearly thirty ATCSs to list tasks that might require them to be able to distinguish and/or name colors (4). Their lists were not expected to be exhaustive, for equipment changes occur with advances in technology. These subject-matter experts, representatives of the major types of air traffic control facilities, listed thirteen tasks which required color vision. The tasks did not require color vision capability of the same kind or degree, and the tasks themselves varied in importance. "Normal color vision" was required to accomplish some of the tasks, and all controllers are expected to be able to perform all of the tasks successfully.

What are some examples of such tasks? When radio communications are not available, a controller may use a red-green light to signal a pilot that he is clear to land. Color can be the means used to identify certain aircraft and their direction of travel, to identify storm centers on a color weather radar screen, and to determine ground elevations on a navigational chart. Continuing changes in equipment generally seem to increase the diversity of tasks to be performed and the number of tasks requiring color perception. When tasks must be performed under less than optimum environmental conditions and under time stress, there will be an increased probability that errors will be made. Errors in air traffic control can have catastrophic consequences with loss of life and property damage. These examples from the OPM analysis (4) illustrate why qualification standards for the ATCS Series GS-2152 state that applicants for initial appointment to this job series must demonstrate normal color vision.

Designing Tests to Measure Performance on ATC Tasks

Example ATCS tasks were selected from the OPM study (2) of ATCS color vision requirements for use in this study. However, this study requires persons who have differences in the ability to perceive colors, and color-defective persons generally are not available in the ATCS workforce since they are excluded during initial applicant screening. Therefore, performance requirements on the problems were structured so they would be appropriate for measuring achievement of persons not trained as air traffic controllers. A test protocol and simulations of ATC tasks were created in three content areas: (1) aircraft colors for fuselage and lights, (2) color weather radar displays, and (3) navigational chart terrain elevations. Two subtests were developed for each content area, resulting in six subtests. The subtests are:

1. Aircraft Colors, Fuselage (30 items, 3 minutes). Part 1 (15 items): identify which of four words (colors) describes the color of a pictured airplane. Part 2 (15 items): identify which of four airplanes was a particular color (colors given in a scale format).

2. Aircraft Colors, Lights (24 items, 3 minutes): identify which of four airplanes is coming toward or going away from examinee by wingtip light combinations (right wingtip light is aviation green, left is red). Sky, city, and rural background distractions are used.

3. Color Weather Radar, Colors (29 items, 3 minutes). Part 1 (15 items): select one of seven color chips that matches the name (four items were not scored because of the peculiar shade of orange used). Part 2 (14 items): select one from seven names to match the color chip that is presented (again four items were not scored because of the peculiar shade of color used).

4. Color Weather Radar Displays (25 items, 3 minutes). Five weather maps with color presentations of storms are presented. The task is to determine weather intensity at numbered locations by comparing locations color to those in a Storm Intensity Code. Colors were obtained from the manufacturer of these displays.

5. Navigational Chart Terrain Elevations Colors (16 items 2 minutes). Part 1 (8 items): match a color to elevations presented in order of actual magnitude. Part 2 (8 items): match colors to elevations with both colors and elevations presented in scrambled order.

6. Navigational Chart Terrain Elevations, Charts (29 items 2 minutes). Four clips from actual navigational charts with color presentations of changing ground elevations are presented. The task is to determine altitude at numbered locations.

Conversion of a draft copy of this six part ATCS task test to a single loose leaf notebook copy suitable for testing purposes was accomplished at a cost approximating $3500. The vendor's response to a requested "time and cost" estimate for future copies was "thirty days and $3500" per copy.

Persons Tested

Color blind persons are rejected when they initially apply for a job as an air traffic controller. Thus it was necessary to search elsewhere for persons with defective color vision. The 63 persons used in the study were all recruited in the Washington, D.C. area, and included United States Coast Guard personnel, FAA employees, graduate students and university professors. Most of the 22 color defectives were Coast Guard headquarters people identified through a search of their medical files, persons administered the PIP at the time of initial entry and retained because color perception capability was not a screening element for their yeoman, storekeeper, or other specialty. Study participants ranged in age from 19 to 66 years, and included 49 males and 14 females. This was a restricted sample in kind as well as size. Normal color vision occurs in about 90 percent of men and 99 percent of women, so a sample representing the civil population of normals and color-defectives would approximate a 9 or 10 to 1 ratio.

Data Collection

Test administrators were trained, and the materials and motivational materials describing the purpose of the study were presented to each person prior to his or her testing. The pseudoisochromatic plate test took about 5 minutes, and the ATCS task simulation testing required about 35 minutes. All persons were administered both tests. Four criterion groups were established: a color perception (1) normal and (2) defective group based on the NAS and FAA Airman First-Class criterion for the plate test of 10 plates or more answered correctly, and a color vision (3) acceptable and (4) unacceptable group based on the FAA Airman Second-Class certificate's less demanding requirement of 5 plates or more answered correctly. Twenty-two persons were classified as color defective using the normal color vision requirement; only eight performances were classified as unacceptable using the FAA's Second-Class certificate requirement.

Results

PIP Plate Difficulties and Test Reliability: Table 1 contains a summary of the difficulties for each plate (proportion of persons correctly answering the item) for (1) the entire group, (2) those classified as color vision normal and (3) defective by the NAS, OPM, FAA First-Class medical certificate criterion, and (4) those classified as color vision unacceptable using the FAA Second-Class certificate definition. Excellent discrimination occurred on all but two plates (8 and 13) between those classified as color vision normal and those classified as defectives and unacceptable. For each criterion, plates 3, 4, 6, 9, 10, 11, and 15 were particularly good discriminators between the color vision normal and unacceptable groups and, except for Plate 10, very easy for the color vision normal groups. The internal consistency reliability of the pseudoisochromatic plates was estimated using a split-half (odd-even) coefficient. An obtained value of .94 compares favorably with the test-retest reliability (using Kappa) coefficient of .96 reported by Seefelt (9).

ATCS Task Test Difficulties and Reliability: All subtests in the simulation of ATC tasks were administered under speeded conditions, as noted in the test descriptions. Thus the number of persons attempting each item gets smaller as testing progresses, and the item subsequently is considered more difficult. Table 2 presents the item difficulties (proportion of persons correctly answering each item) for the six ATCS task subtests. The effects of the speeded conditions are notable in all instances. Subtest 5, Navigational Chart Colors, was particularly difficult. The task is very challenging, and it is unlikely that the performance of the group can be explained solely from the effects of the speeded conditions. Table 3 contains the split-half (odd-even) reliability estimate for each subtest. Normally, reliability coefficients greater than or equal to .80 are desired. Only the chart portion of the Navigational Chart Terrain Elevations Test failed to meet this goal. A reliability estimate for the composite yielded a value of .96.

<u>PIP Test Validities, Ten plates or more correct</u>: Table 4 presents the means and standard deviations of the number of items answered correctly (RIGHT), answered incorrectly (WRONG), and omitted (OMIT) for each ATCS task for the color vision normal (10 plates or more correct) and defective group (less than 10 plates correct) on the PIP. In addition, the t-value and its associated probability level resulting from testing the difference between the RIGHT scores for the two groups are given in parenthesis for each subtest. Four of the t-values exceed the .01 level, one meets the .01 level, and one is at the .03 level. Thus the PIP normal color vision criterion demonstrates a relation with these ATCS tasks that meets and exceeds the Uniform Guidelines' .05 level requirement. The multiple correlation for the six ATCS tasks with the color vision normal criterion (Table 7) yielded a validity coefficient of .565.

<u>PIP Test Validities, 5 Plates or more correct</u>: Table 5 presents for the satisfactory and unsatisfactory color vision groups as determined by the FAA Second-Class Certificate criterion, the same type of information as in Table 4. As expected, the mean RIGHT score for the unsatisfactory color vision group in Table 5 was less than the mean RIGHT score for the color vision defective group in Table 4, except for the Aircraft Lights subtest. However, within Table 5, the t-value and its associated probability level resulting from testing the difference between the RIGHT scores for the satisfactory and unsatisfactory groups are given for each subtest, and these meet the Uniform Guidelines .05 criterion for five of the six subtests. Only the Navigational Chart Terrain Elevations Color Test fails to meet that requirement. The multiple correlation for the six ATCS tasks with this FAA Second-Class certificate satisfactory-unsatisfactory definition (Table 7) yielded a validity coefficient of .490.

<u>Correlational Analysis</u>: Table 6 contains the correlations between the scores from all ATCS Task subtests, the PIP raw scores, and the groups determined by the PIP normal color vision criterion score (COLOR 10) and by the FAA Second-Class certificate criterion score (COLOR 5). The following trends are evident from an inspection of Table 6:

- Five of the six correlations between the PIP raw scores and the ATCS Task subtests ranged from .43 to .48; the remaining correlation with the Terrain Elevation Color Test was .26. These correlations are moderate and typical of predictive validity coefficients found in the behavioral sciences.

- Five of the six correlations between the classification categories established with the normal color vision score (COLOR 10) and the tests were slightly lower than the corresponding correlations between the PIP raw scores and the tests. The average decrease in the corresponding correlations was .046. This decrease is due to the change in the shape of the distribution when a continuous variable (PIP raw score) is transformed to a dichotomous variable (COLOR 10).

- Five of the six correlations between the classification categories using the FAA Second-Class certificate criterion (COLOR 5) and the tests were even lower than the corresponding correlations between COLOR 10 and the tests.

- The intercorrelations among the subtests ranged from .31 to .72. The average correlations in descending order are: AIRCRAFT COLORS (.552), RADAR COLORS (.544), RADAR DISPLAY (.538), TERRAIN COLORS (.470), TERRAIN CHART (.422), and AIRCRAFT LIGHTS (.392). The subtests AIRCRAFT COLORS, RADAR COLORS, and RADAR DISPLAY form a rather distinct cluster with TERRAIN COLORS being a possible member of that cluster.

In addition, a score corrected for guessing (4 RIGHT - WRONG) was created for each test. Correlations were computed between the tests using the corrected score, the tests using the original RIGHT score, and PIP, COLOR 10 and COLOR 5. This resulted in very little change from the correlations presented in Table 6, primarily because the corrected score was virtually a linear transformation of the RIGHT score (if every examinee answered the same number of items, the transformation would be linear). The subtest correlations between corrected and original scores were either .98 or .99. However, multiple correlations for the six corrected ATCS task scores with the PIP, COLOR 10 and COLOR 5, increased to .580 and .505 respectively (Table 7).

<u>Summary</u>

On the basis of the above analyses, moderate to good support exists for application of the PIP test as a screening device for ATCS tasks involving color discriminations. Such findings are specific to a PIP test as used in this study, and they support application of either the FAA Airman Second-Class or Airman First-Class criterion. However, the evidence in this study and in the research literature is more substantial for continued application of the OPM "normal color vision" standard.

DEVELOPMENT OF FUNCTIONAL COLOR VISION TESTS

The OPM had recommended (4) development of functional color vision tests "to reflect as closely as possible the functional color vision requirements of the specialty," and the ATC task tests developed here for validation of the PIP fit that description. Actual navigational chart segments are used in one of the tests, with the task of using chart coloring to determine altitudes at numbered locations. In another, pictures of color weather radar scopes present weather maps in colors provided by the manufacturer of those scopes, and the task is to determine weather intensities at numbered locations. In a third task the person is to determine which of four aircraft is approaching or departing by wingtip light combinations (right wingtip light is aviation green, and the left is red). The difficulty level of these tasks was designed for accomplishment by persons not trained as air traffic controllers, and their success is demonstrated for the persons tested in this study. Reliability estimates for the subtests and composite are satisfactory.

The issue of a proper statistical validation study remains, but the OPM has validated the color vision task need and a true-on-the-job validity study is not feasible. These tests all have high content validity. The validity of the functional tests for use as measures of color vision should be demonstrated by relating such scores to performance on the anomaloscope, but that is beyond the scope of this study. Since scores on the PIP correlate highly with performance on the anomaloscope, the PIP test scores were used as a substitute, for study of these functional tests and their expected validity for identifying color defective persons. Regression analyses were performed using the functional tests to predict PIP scores, and discriminant function and classification analyses to study the false positives and false negatives identified in the prediction process.

Regression Analysis

Two multiple linear regression analyses were performed for each of three dependent variables, resulting in a total of six analyses. The dependent variables were the PIP raw scores, the COLOR 10 classification, and the COLOR 5 classification. For each dependent variable, two sets of six predictors were used as follows:

(1) subtest totals: the RIGHT score for each of the ATC task simulation subtests; i.e. Aircraft Colors Fuselage & Lights, Weather Radar Colors & Display, and Terrain Elevation Colors & Chart;

(2) corrected totals: the total subtest score corrected for guessing (4 RIGHT - WRONG) for each subtest.

Inspection of the results of these analyses, given in Table 7, reveals the following:

- For each dependent variable, the highest multiple correlation (R) occurs when the corrected totals are used, and the lower multiple R occurs when the subtest totals are used.
- Better prediction of defective color perception occurs as expected when the continuous dependent variable (PIP Raw Score) is used than when either of the dichotomous variables, COLOR 10 or COLOR 5, IS USED.
- The prediction of defective color perception using COLOR 10 is better than the prediction using COLOR 5.

- The multiple R's ranged from a high of .628 (PIP Raw Score with the Corrected Total) to a low of .490 (COLOR 5 with the subtest totals as predictors). The variance explained ranged from 42.7% of the total variance to 24.0% of the total variance. The multiple R's are restricted somewhat by the moderate level of multicollinearity (inter- correlation among the predictors) in the predictor set.

- The relative importance of each predictor, as determined by its regression weight, changes according to which dependent variable and which set of predictors are used. The patterns are as follows:

    1. With PIP Raw Score and COLOR 10 as dependent:

        (a) generally Aircraft Lights, Radar Colors, and Terrain Chart are the most important.

        (b) Terrain Colors and Radar Display are either very unimportant (near zero weight) or have negative weights.

    2. With COLOR 5 as dependent:

        (a) Radar Display and Terrain Chart are consistently important predictors with Radar Display being more important than Terrain Chart.

        (b) Aircraft Lights, Radar Colors, and Terrain Colors consistently have negative weights.

Discriminant Function Analysis

Two separate discriminant function analyses were performed on the data, one using the groups established by the criterion of the usual score (COLOR 10) and the other using the groups established by the FAA Second-Class Certificate score (COLOR 5). The results of the analyses are given in Table 8. As expected, the results of the discriminant function analyses parallel the results of the regression analyses since only two groups, color perception acceptable and defective, are present in each analysis. A classification analysis was performed based on the results of each discriminant function analysis. The analysis compares the number of examinees predicted to be color perception acceptable and defective on the basis of the ATCS task subtests with the number determined to be color perception acceptable and defective on the basis of the PIP Color 10 and Color 5 scores. In the classification analysis, three statistics are of interest:

(1) the percentage of examinees correctly classified on the basis of the subtest performance. This ranged from a high of 82.54% for COLOR 5 using the subtest totals to a low of 76.19% for COLOR 10 using the sub- test totals.

(2) the percentage of false positives, those examinees predicted to be color perception defective on the basis of the ATCS task subtests who were classified as color vision normal by the PIP. This percentage ranged from a high of 17.46% for COLOR 5 with the subtest totals to a low of 11.11% for COLOR 10 with the subtest totals.

(3) the percentage of false negatives, those examinees predicted to be color perception acceptable on the basis of the ATCS task subtests who were classified as color defective by the PIP. This percentage ranged from a high of 9.52% to a low of 3.17% of the total number of examinees.

Assuming the above designations of false positive and false negative are meaningful, then from a public perspective (i.e., safety) a false positive is a far more serious error than a false negative. When viewed from the perspective of the individual (i.e., fairness in hiring), the seriousness of the false negative increases considerably.

Summary

The functional color vision tests developed here have high content validity for ATC tasks and demonstrate potentially high validity for correctly classifying persons of known varying color perception capabilities. Proper statistical studies validating them against on-the-job performance and validating them against anomaloscope scores would be desirable. Creation of a total score for the three test composite and pass-fail cutoff scores would be needed if the test is to be used operationally. From a cost analysis point of view, most AMEs already have copy of a PIP, and new copies are available at a cost of approximately $62 each. Only a single copy of the functional test is available, and new copies are estimated to cost approximately $3500 each.

FINDINGS AND CONCLUSIONS

On the basis of the above information, moderate to good support exists for use of the PIP as a screening device for ATC tasks involving color discriminations. An extensive body of scientific literature is available as evidence of its reliability and validity for measurement of genetic color perception deficiencies. Evidence of its reliability also was obtained in this study, plus evidence of its validity as a screening device for ATC tasks involving color discriminations.

The Uniform Guidelines requirement for a demonstrated relation between performance on-the-job selection procedure (PIP) and performance on ATC tasks that is statistically significant at the .05 probability level, is met. The findings in this study support application of either the FAA Airman Second-Class or Airman First-Class standard, but the evidence here and in the research literature is more substantial for continued application of the OPM "normal color vision" standard.

The functional color vision tests developed here for validating the PIP test have high content validity for ATC tasks and demonstrate potentially high validity for use in more detailed examination of those identified as color defective. These tests demonstrated the desirable technical properties of satisfactory reliability and reasonable item difficulties. Other supportive information includes moderate correlations (a few exceeding .60 might be classified as high) among these simulated ATC tasks and between PIP scores and task performances. These correlations were typical of what is found in the behavioral science literature. Moderately high and typical R and R^2 values resulted from the regressions of PIP raw scores, normal color vision (10 plates or more correct), and FAA Airman Second-Class (5 plates or more correct) standards on representative predictors from the ATC tasks. A high percentage of correct color perception normal and color defective classifications also were obtained, based on the discriminant function analyses.

Further study of these functional color vision tests is desirable before their operational use should be considered. Supportive validation of the PIP for ATC task performances has been provided, and those tests are commonly used by AMEs for testing ATCS applicants. Their cost is approximately $62 per copy; estimated cost per copy of the functional test designed for this study is $3500. Validation of the new test against on-the-job performance and performance on the anomaloscope would be desirable. Repeat experimental studies, creation of a total score for the three test composite, and pass-fail cut off scores for operational application, would be needed.

Continued use of the PIP, with standardized administration and application of the OPM "normal color vision" standard is supported.

Table 1

Pseudoisochromatic Plates

Proportion of Persons Answering the Plate Correctly

Plate Number	Entire Group	Color Vision Groups		
		Normal[1]	Defective[1]	Unacceptable[2]
	(n = 63)	(n=41)	(n=22)	(n=8)
2	.87	1.00	.64	.38
3	.65	.98	.05	.00
4	.70	1.00	.14	.00
5	.89	1.00	.68	.50
6	.57	.90	.05	.00
7	.89	1.00	.68	.38
8	.92	.98	.82	.63
9	.54	.76	.14	.00
10	.27	.41	.00	.00
11	.56	.85	.00	.00
12	.84	.98	.50	.25
13	.95	1.00	.86	.88
14	.86	.98	.64	.38
15	.60	.90	.05	.00

[1] AOC PIP test definition

[2] FAA Second-Class Medical Certificate

Table 2

ATCS Tasks

Proportion of Persons Answering Items Correctly

Subtest	Aircraft		Color Weather Radar		Nav. Chart	Terrain Elev.
Item	Fuselage	Lights	Colors (a)	Display	Colors	Chart
1	.79	.71	1.00	.87	.59	.75
2	.49	.68	-	.73	.59	.68
3	.75	.67	-	.94	.54	.11
4	.87	.68	.95	.75	.78	.56
5	.94	.68	-	.62	.71	.62
6	.81	.68	.98	.79	.71	.19
7	.94	.63	.92	.94	.79	.54
8	.84	.60	.78	.87	.84	.71
9	.56	.65	.95	.81	.16	.65
10	.68	.59	.94	.95	.13	.65
11	.94	.70	.87	.98	.27	.71
12	.32	.65	.92	.89	.19	.75
13	.87	.56	-	.84	.02	.43
14	.76	.59	.97	.75	.06	.32
15	.81	.56	.92	.83	.06	.71
16	.79	.54	-	.73	.16	.38
17	.71	.46	-	.75		.79
18	.83	.41	.65	.70		.56
19	.79	.33	.81	.62		.75
20	.71	.27	.79	.60		.41
21	.57	.05	-	.48		.27
22	.05	.14	.86	.33		.29
23	.14	.11	.73	.37		.43
24	.43	.02	-	.24		.43
25	.37		.56	.27		.51
26	.27		.41			.46
27	.19		.43			.51
28	.16		.40			.33
29	.21		.40			.08
30	.14					

a. Items 2, 3, 5, 13, 16, 17, 21, and 24 were not scored because of an unusual shade of orange used.

Table 3

Simulated ATCS Tasks

Split-Half (Odd-Even) Reliability Estimates

Test	Estimate
Aircraft Colors	.98
1. Fuselage	.97
2. Lights	.97
Weather Radar	.96
3. Colors	.94
4. Display	.94
Navigational Chart Terrain Elevation	.84
5. Colors	.89
6. Chart	.70
Composite (Total, all subtests)	.96

Table 4

ATCS Task Performances of Criterion Groups as Determined by Normal Color Vision Criterion*

ATCS Task Subtests		RIGHT Defective	RIGHT Normal	t	WRONG Defective	WRONG Normal	OMIT Defective	OMIT Normal
Aircraft Colors								
1. Fuselage	MEAN	14.09	19.63	3.77	7.68	4.02	8.23	6.34
	STD DEV	6.47	5.01	(000)	4.04	3.83	4.26	4.64
2. Lights	MEAN	8.32	14.54	3.58	6.77	4.99	8.91	6.02
	STD DEV	6.61	6.56	(001)	3.44	4.79	3.75	3.92
Weather Radar								
3. Colors	MEAN	13.82	17.51	3.30	2.45	1.07	4.73	2.41
	STD DEV	4.76	3.03	(002)	2.54	1.25	3.41	2.64
4. Display	MEAN	14.95	19.10	2.64	3.50	1.98	6.55	3.93
	STD DEV	6.57	4.51	(013)	3.35	1.97	4.52	4.41
Nav. Chart Terrain Elev.								
5. Colors	MEAN	5.32	7.02	2.17	4.77	2.71	5.91	6.27
	STD DEV	2.68	3.11	(034)	2.29	2.37	2.67	2.51
6. Chart	MEAN	12.45	15.85	3.37	14.23	10.51	2.32	2.63
	STD DEV	2.92	4.21	(001)	2.72	4.77	3.80	3.90

*PIP Scoring Criterion: Defective - less than 10 correct (n = 22)
Normal - 10 or more correct (n=41)

Table 5

ATCS Task Performances of Criterion Groups as Determined by FAA's Second Class Requirement, PIP Score[a]

ATCS Task Subtests		RIGHT Unacceptable	RIGHT Acceptable	t	WRONG Unacceptable	WRONG Acceptable	OMIT Unacceptable	OMIT Acceptable
Aircraft Colors								
1. Fuselage	MEAN	12.50	18.45	2.70	8.00	4.91	9.50	6.64
	STD DEV	6.14	5.79	(009)	3.55	4.23	4.72	4.45
2. Lights	MEAN	9.00	12.85	2.38	6.63	4.31	8.38	6.84
	STD DEV	3.59	7.45	(028)	4.50	5.13	3.66	4.13
Weather Radar								
3. Colors	MEAN	13.13	16.67	2.38	2.75	1.38	5.13	2.95
	STD DEV	5.54	3.69	(021)	2.77	1.72	4.02	2.90
4. Display	MEAN	11.25	18.58	3.80	5.75	2.04	8.00	4.38
	STD DEV	7.17	4.77	(000)	3.20	2.18	4.96	4.13
Nav. Chart Terrain Elev.								
5. Colors	MEAN	5.00	6.64	1.43	4.38	3.29	6.63	6.07
	STD DEV	2.67	3.08	(159)	2.07	2.28	2.13	2.62
6. Chart	MEAN	11.00	15.20	2.84	14.88	11.36	3.13	2.44
	STD DEV	3.02	4.01	(006)	4.19	4.20	5.19	3.66

[a]Scoring Criterion: Unacceptable - less than 5 correct (n = 8)
Acceptable - 5 or more correct (n = 55)

Table 6

Correlation Matrix for RIGHT Score

ATCS Task Subtests and AOC Plates

ATCS Task Subtests Variable	1	2	3	4	5	6	7	8
Aircraft Colors								
1. Fuselage	-							
2. Lights	52	-						
Weather Radar								
3. Colors	67	33	-					
4. Display	66	44	72	-				
Nav. Chart Terrain Elev.								
5. Colors	50	33	55	45	-			
6. Chart	41	31	45	42	52	-		
AOC Pseudoisochromatic Plates								
7. Raw Score	48	43	47	46	26	43	-	
8. Normal Color Vision (COLOR 10)	44	42	43	35	27	40	91	-
9. FAA Second-Class (COLOR 5)	33	18	29	44	18	34	72	52
	1	2	3	4	5	6	7	8

Table 7

Regression Summaries

				ATCS Task Subtests					
				Aircraft Colors		Weather Radar		Terrain Elevations	
Dependent	Predictors	R	R^2	Fuselage	Lights	Colors	Display	Colors	Charts
COLOR 10	Subtest totals	.565	.320	.14	.26	.29	-.10	-.13	.24
COLOR 10	Corrected total	.580	.337	.18	.24	.25	-.10	-.11	.28
COLOR 5	Subtest totals	.490	.240	.11	-.06	-.11	.42	-.11	.29
COLOR 5	Corrected total	.505	.255	.12	-.05	-.11	.42	-.12	.25
PIP Raw Score	Subtest totals	.614	.378	.15	.23	.23	.06	-.19	.27
PIP Raw Score	Corrected total	.628	.394	.18	.22	.19	.08	-.17	.30

Table 8

Discriminant Function and Classification Results

Coefficients	Groups: COLOR 10 (Subtest totals)	Groups: COLOR 5 (Subtest totals)
Aircraft Colors		
1. Fuselage	.26	.25
2. Lights	.51	-.14
Weather Radar		
3. Colors	.57	-.25
4. Display	-.21	.87
Nav. Chart Terrain Elev.		
5. Colors	-.27	-.25
6. Chart	.47	.55

		Predicted: ATCS Classification Def	ATCS Classification Norm	Color 10 Def	Color 10 Norm	ATCS Classification Def	ATCS Classification Norm	Color 5 Def	Color 5 Norm
Actual (PIP)	Defective	17	5	16	6	6	2	6	2
	Normal	10	31	7	34	9	46	11	44
	Correctly Classified	76.19%		79.37%		82.54%		79.37%	

REFERENCES

1. Qualification Standards, Air Traffic Controller Series, U.S. Office of Personnel Management Handbook X-118, Part III, 1980.

2. Guide for Aviation Medical Examiners, Federal Aviation Administration, U.S. Department of Transportation, 800 Independence Ave., SW, Washington, D.C., October 1981.

3. Federal Personnel Manual 339, Supplement 339-31, and Letter 339-15, U.S. Office of Personnel Management, Washington, D.C.

4. Unpublished memorandum from Helen Christrup, Assistant Director for Personnel Research and Development, OPM, to Donald B. Rock, Director of Personnel and Training, FAA, dated October 16, 1981.

5. "Color Blindness and Abnormalities of Vision," D. B. Judd, in Contributions to Color Science, National Bureau of Standards Special Publication 545, U. S. Department of Commerce, September 1979.

6. Federal Register, Vol. 43, No. 166, August 25, 1978, pages 38295-38310

7. Procedures for Testing Color Vision, Committee on Vision, Assembly of Behavioral and Social Sciences, National Research Council, National Accademy of Sciences, National Academy Press, Washington, D.C. 1981.

8. Pseudoisochromatic Plates for Testing Color Perception. American Optical Corporation, Copywrite 1965

9. Seefelt, E. R. An evaluation of the validity and reliability of the AOC 15-plate pseudoisochromatic test in routine testing. American Journal of Optometry, 1964, 41, 371-381.