DR. T. EUGENE AVERY*
University of Illinois
Urbana, Ill. 61801
HAROLD BURKHART†
University of Georgia
Athens, Georgia 30601

Screening Tests for Rating Photo Interpreters

The reliability of correct selection can probably be increased from about 41% (without testing) to about 73%.

(*Abstract on page* 478)

THE PROBLEM

IF ASKED TO LIST the essential qualifications for persons to be trained as photo interpreters, one might logically include good stereoscopic vision, normal color perception, a sound educational background, and above-average mental ability or I.Q. rating. But would the use of these criteria alone consistently separate good potential photo interpreters from poor ones? Probably not! Thus, to supplement the accepted testing procedures for evaluating these qualities, we also need some measure of an individual's motivation, judgment, and logical reasoning abilities. To state the problem in another way, can we develop a screening process for selecting the 50 best potential interpreters from an apparently homogeneous group of 300 or more applicants?

The need for a battery of screening tests for rating trainees *prior* to specialized photo interpretation (P. I.) instruction was outlined in 1956 by Sims and Hall. In 1960, Meyer and Miller reported that photo interpretation performance, based on recognition and assessment of industrial features, could be predicted with reasonable success from scores on aptitude tests (problem solving and spatial relations) and visual perception examinations. Following a similar study, it was concluded by Sadacca, Martinek, and Schwartz (1962) that the task of image interpretation is highly cognitive, and if basic visual standards are met, the ability to make necessary complex decisions and judgments will depend mostly on the mental capacity, interest, and experience of the individual.

In 1965, Avery reported on the initiation of a research study aimed at determining (1) whether the elusive qualities of motivation, judgment, imagination, and logical reasoning could be determined by various screening tests, and (2) whether scores on such screening tests could be correlated with interpretation performance at the termination of a specialized P.I. training course. This report summarizes the results of that study.

DR. T. EUGENE AVERY

* Head, Department of Forestry.
† Doctoral candidate in the School of Forestry.

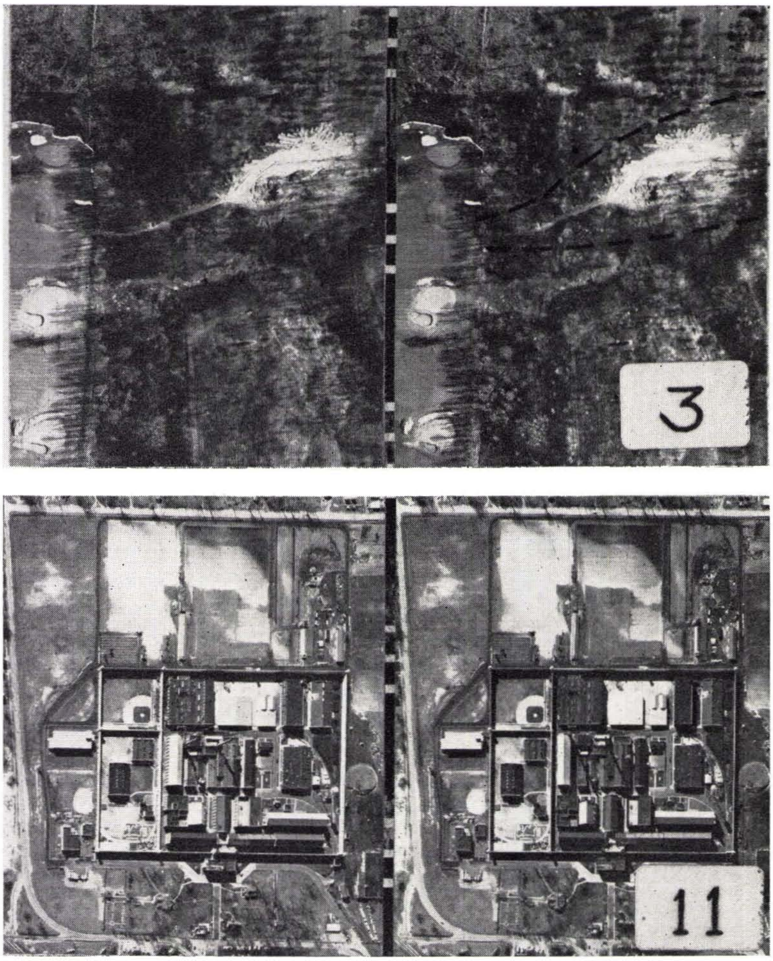

FIG. 1. Examples of stereograms from the Stereo Performance Test. The numbers "3" and "11" correspond to the questions and answers listed in the text of this article.

SCREENING TESTS

A total of 99 male college students at the University of Georgia participated as trainees in the testing experiment. Most of the men were 20 to 25 years old, and all had good stereoscopic vision. These students were enrolled in beginning classes in photo interpretation during the 1964–65 and 1965–66 academic years; none had received previous civilian or military training in aerial photo interpretation.* At the beginning of classes, following a brief introduction to stereoscopic viewing, a battery of seven screening tests was administered to each trainee. The tests were given in the following order:

1. *Simulated photographic images test* (Avery, 1964). This test requires identification of 39 sketches that depict overhead views or shadows of common urban, rural, or physiographic features.
2. *Aerial photo identification test* (Avery, 1964). Twenty-four common urban and rural features must be identified on vertical photographs.
3. *Picture integration puzzle* (Avery, 1964). This type of test requires that examinees assemble cut-out portions of an aerial photograph into a composite and complete picture. Scoring is based on the time required to complete the puzzle.
4. *Mental ability or I.Q. test* (Otis, 1954). The Otis test is a group-administered intelligence quotient exam consisting of 80 questions on vocabulary, word relationships, progressions, mathematics, and figure analogies.
5. *Object visualization test* (Miller, 1955). This test consists of 44 drawings; geometric

* To obtain the highest possible degree of homogeneity in the trainee group, results of female students were not included in the analysis. Similarly, students with poor stereoscopic acuity and those with previous P.I. training were excluded. Participation in the experiment was voluntary, and results had no bearing on an individual's final college grade in the course. The course included about 75 hours of classroom and laboratory instruction in elementary photogrammetry, photo interpretation, and mapping techniques.

solids are presented as flat or unfolded patterns. Examinees must visualize the geometric shapes that will be formed when the flat patterns are correctly folded or rolled together.

6. *Standard progressive matrices* (Raven, 1958). Sixty figure analogies are utilized to measure an individual's capacity for observation, clear thinking, and logical reasoning.

7. *Stereo practice quiz* (Avery, 1964). Consisting of 25 stereograms, this unpublished test requires the student to make image identifications, object counts, height rankings, and one length measurement.

Except for the picture integration puzzle and mental ability test, scores were recorded as a percent of correct responses. Scoring of the picture integration puzzle test was based on the time required to complete the puzzle and computed as 1,000 divided by the puzzle assembly time in minutes. Mental ability was scored as the intelligence quotient indicated by the Otis test.

ABSTRACT: *Can individuals be rated according to photo interpretation ability prior to specialized training? Results of a two-year study involving 99 university students indicate that some measure of photo interpretation performance may be reliably predicted by scores on two specialized screening tests administered at the onset of training. One required the identification of 39 sketches that depict overhead views of common urban, rural, and physiographic features. The other test consisted of 25 image identifications and measurements made on prepared stereograms. In combination, scores on these screening examinations accounted for about one-hc lf of the variability in final photo interpretation performance ratings.*

The Elusive Measure of Performance

To determine the relative values of the screening tests administered at the onset of training, an objective measure of performance was required at the end of the photo interpretation course. The performance criteria most frequently employed in the other studies investigated was the composite course grade or overall proficiency record of each individual at the termination of a training period. Scholastic grades were deemed unsuitable in this study, however, because these composite scores were based on written examinations and map compilations as well as on photo interpretation abilities. What was desired was a final performance rating based on visual scanning ability and on the reliability of photo identifications. Devising such performance tests was complicated by two noncontrollable elements: the inevitable differences in cultural backgrounds of the 99 trainees, and the fact that the students were majoring in such diverse subject-matter fields as geography, geology, forestry, landscape architecture, and physics.

In spite of the foregoing difficulties, two performance tests were devised. The first, a *Visual Search Test*, is designed to evaluate photo scanning ability. Sometimes referred to as a spatial orientation or grid location test, this examination requires the student to determine the exact coordinate positions of small photo cut-outs on a complete vertical print of the same area. In addition, one part of the test requires the trainee to make correct translations from map locations to an aerial photo index of the same area. Examinees were rated on a percentage scale according to the time required to list the correct grid coordinates on a special answer form.

The second part of the performance rating was based on identifications of cultural features or *signposts of civilization* as depicted on 25 specially prepared vertical stereograms. Insofar as available photographs permitted, the test items were chosen to evaluate an individual's abilities in deductive reasoning rather than requiring mere recognition by rote or memorization. Because of the particular associations of features pictured, this examination proved much more difficult than the analogous *stereo-practice quiz* used as a screening test. Nevertheless, the vertical views illustrated features which students would have quickly recognized from close ground observation; the less familiar overhead views thus served to separate the thinking interpreters from their less perceptive classmates.

A single composite performance rating for each student was derived from a weighted average of scores received on the *Visual Search Test* (weight 1/4) and the *Stereo Per-*

formance Test (weight 3/4). It seemed that these relative weights for photo scanning ability and image identification provided a realistic measure of the type of performance often expected of both military and civilian interpreters.

Analysis of Testing Results

As outlined earlier, the primary objective of this study was to determine whether scores on one or more of the preliminary screening tests could be used to predict the final performance rating of trainees. Accordingly, multiple regression techniques were employed to analyze relationships between the seven screening test scores and the final performance rating for each of the 99 participating individuals.

The data were analyzed at the University of Georgia Computer Center using a Biomedical computer program designated as BMDO2R. This program carries out the regression analysis in a stepwise fashion as described by Efroymsen (1960). In addition to the regression analysis, the program provides the mean and standard deviation for each of the independent and dependent variables, a covariance matrix, and a correlation matrix. Plottings of residual sums of squares on each independent variable are also carried out. In all cases, there was a random scatter of points about the zero line, further supporting the validity of a linear statistical model assumption.

The analysis indicated that the Simulated Photo Images Test (Screening Test No. 1) was the best single predictor of performance, and it accounted for 38 percent of the variability in performance scores. The Stereo Practice Quiz (Screening Test No. 7) was removed on the second step, and, combined with the Simulated Photo Images Test score, boosted the amount of performance variability accounted for to 48 percent. The Standard Progressive Matrices Test was the third independent variable removed, but this only increased the predictive ability by about 2 percent. Consequently, this test was omitted from the prediction equation. The remaining screening tests, when entered in the prediction equation, each added 1 percent or less in accounting for the total amount of variability in performance scores.

A stepwise regression analysis using grades from the Simulated Photo Images Test and the Stereo Performance Quiz yielded the following prediction equation:

Predicted Performance Rating $= 19.16108 + 0.41728\ X_1 + 0.36619\ X_2$,

Table 1.

Screening test	*Correlation coefficient with performance rating*
Simulated Photo Images	0.62**
Stereo Practice Quiz	0.60**

** Significant at the .01 level.

where

X_1 = Simulated Photo Image Test Score

X_2 = Stereo Practice Quiz Score

When applying this prediction equation, the independent variables (X_1 and X_2) must be entered as percent of correct responses on the respective tests. This will provide a predicted performance rating in percent. Simple linear correlation coefficients between the two screening tests used in the prediction equation and the final performance rating are shown in the tabulation that follows. These correlation coefficients substantiate the highly significant association between screening test scores and final performance ratings (Table 1).

Discussion and Conclusions

Assuming that the proposed method of rating interpreter performance is valid, the more promising candidates for P.I. training may be selected on the basis of their scores on two screening tests. The prediction equation accounts for 48 percent of the variability that will exist in a trainee's final performance scores. Thus, the equation permits the ranking of individuals as an aid in selection for training. With their predicted performance scores arranged in rank order, the required number of P.I. trainees could be chosen by starting with the highest ranking individual and working downward. Accuracy of ranking will depend on the size of the group and relative aptitudes of individuals making up the group. As a practical example, if a group of 50 P.I. candidates are to be selected for training, it would be desirable to screen them from a group of 300 or more individuals rather than from a group of only 100 persons.

To demonstrate further the validity of the selection criteria proposed here, performance ratings for the 99 students participating in this study were computed by substituting their screening test scores in the prediction equation. The *predicted* performance ratings were then checked against actual performance scores for each man. A performance rating of 80 percent or greater was arbitrarily classed as satisfactory; ratings of 79 percent and lower were deemed unsatisfactory. Compari-

TABLE 2.

Predicted performance rating	Actual performance rating		
	Unsatisfactory	*Satisfactory*	*Totals*
Satisfactory	10	25	35
Unsatisfactory	48	16	64
Totals	58	41	99

sons of predicted and actual ratings were as shown in Table 2.

Of the 35 persons predicted as satisfactory, 25 actually proved to be satisfactory, while 10 individuals had unsatisfactory performance ratings. Furthermore, of the 64 persons predicted to be unsatisfactory, 48 actually proved to be unsatisfactory, although 16 did achieve satisfactory performance ratings. Thus, by regarding 80 percent on performance scores as an acceptable screening level, it would have been possible to categorize correctly 73 of the 99 men tested here as satisfactory or unsatisfactory in advance of P.I. training.

AVAILABILITY OF TESTS

Because of costs involved, it has not been feasible to reproduce here the screening and performance tests employed in this study. However, sample questions from the various screening tests are illustrated in Avery (1965) cited at the end of this report. Copies of the Screening Tests 1 and 2 and the Visual Search Test are available at nominal cost from the senior author.

The Stereo Practice Quiz and the Stereo Performance Test consisted of 25 black-and-white stereograms each, assembled and annotated by the senior author. As these stereograms are in unpublished form, a listing of correct responses for each question is included here to indicate the range of imagery and information required. Both tests utilized panchromatic photographs at a scale of 660 feet per inch. Stereograms were arranged on laboratory tables, and trainees were allowed three minutes per stereogram or a total of 75 minutes for each test (Figures 1, 2).

STEREO PRACTICE QUIZ—REQUIRED IDENTIFICATIONS

1. (a) Gas station (b) church (c) shopping center
2. Contour plowing, or terraced cropland
3. (a) Golf course (b) tennis court (c) swimming pool
4. Drive-in theatre
5. Light manufacturing industry
6. Recognition of cedar tree from shadow pattern
7. Orchard of fruit trees
8. Grade school
9. Graded dirt road
10. (a) School busses (b) football field (c) water tank
11. Bridge construction for main thoroughfare
12. College or university campus
13. (a) Auto junkyard (b) small stream or creek
14. Auto garages for residences
15. (a) steel electrical tower (b) river (c) concrete highway bridge
16. Trucking terminal
17. Cleared rights-of-way
18. Concrete dam
19. Mobile homes
20. Baseball stadium
21. Count types of railroad cars in a train
22. Count number of tracks at railroad grade crossing
23. Rank several building roofs in height order
24. Read name on building roof
25. Determine length of a building

STEREO PERFORMANCE TEST—REQUIRED IDENTIFICATIONS

1. Identify group of cattle in pasture near decaying foundation of old residential building
2. Determine time of day photography from shadow patterns
3. Determine use of sand being hauled from pit (used for sand traps on nearby golf course).
4. Determine reason for land clearing (new housing development).
5. Identify (a) former sites of oil tanks, (b) sand and gravel extraction.
6. Identify private estate for training race horses.
7. (a) Explain pattern (piles of dumped sand from trucks). (b) Identify sand hoppers for loading dump trucks.
8. Identify oil drilling and pumping station.
9. Identify city bus terminal and bus turn-around facility.
10. Deduce meaning of numbers painted on a building roof (longitude and latitude of city).
11. Identify state prison and farm.
12. Identify oil refinery and petroleum storage facilities.
13. Identify cemetery and adjacent city water purification plant.
14. Deduce the principal raw materials (coal, water) and product (electricity) of an industry (electrical power plant).
15. Identify stacks of lumber, poles, and other wood products in an industrial area.
16. Identify and explain function of railroad turntable and roundhouse.
17. Explain purpose of cleared strip through forest cover (electric power transmission line).
18. Identify products being hauled in trucks (petroleum products).

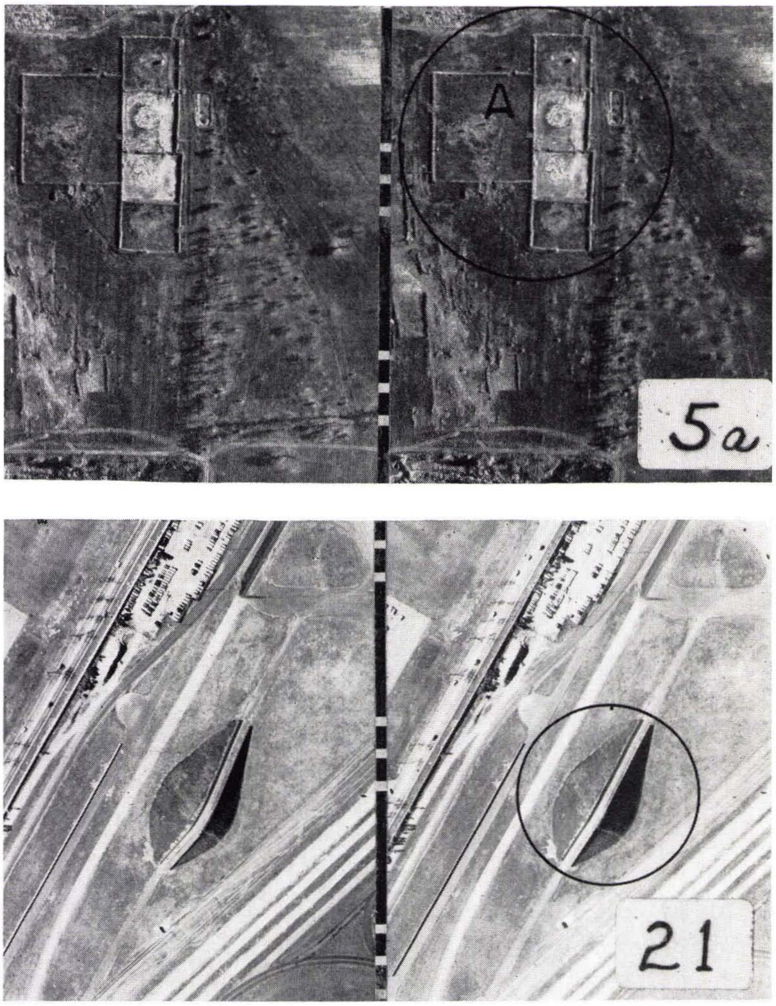

FIG. 2. Examples of stereograms from the Stereo Performance Test. The numbers "5" and "21" correspond to the questions and answers in the text of this article.

19. Identify empty automobile carriers in warehouse parking lot.
20. Explain (a) activity being carried out on ship (refueling or discharge of petroleum) (b) construction material of bridge (steel grating).
21. Identity part of an automobile test rack.
22. Identity steel manufacturing plant from exposed raw materials.
23. Identify hospital and large school in close proximity.
24. Identify (a) railroad passenger terminal (b) railroad bridge construction.
25. Identify a tunnel under street for pedestrians and a covered outside walk.

REFERENCES

Avery, T. Eugene. 1965. Evaluating the potential of photo interpreters. *Photogrammetric Engineering*, 31(6): 1051–1059, illus.

Colwell, Robert N. 1961. Some recent developments of basic significance for photo interpretation. Paper presented at Third Regional Cartographic Conference of the United Nations for Asia and the Far East, Bangkok, Thailand, Oct. 27–Nov. 10, 27 pp., illus.

Dixon, W. J., Editor. 1964. *BMD Biomedical Computer Programs*. Health Sciences Computing Facility, Dept. of Preventative Medicine and Public Health, School of Medicine, Univ. of Cal., Los Angeles, Cal., 585 pp.

Efroymsen, M. A. 1960. *Mathematical Methods for Digital Computers*, multiple regression analysis. Edited by A. Ralston and H. S. Wilf, John Wiley and Sons, Inc., New York, N. Y., pp. 191–203.

Meyer, John K. and Miller, Robert E. 1960. Validity of photo interpreter predictors for test and training criteria. Wright Air Dev. Div., U.S.A.F., Lackland Air Force Base, Texas. Task 7719, Project 17108, 5 pp.

Miller, Daniel R. 1955. Survey of object visualization. California Test Bureau, Monterey, Cal. 4 pp., illus.

Otis, Arthur S. 1954. Otis quick-scoring mental ability test: Gamma test, form FM. Harcourt, Brace and World, Inc., New York, N. Y., 6 pp.

Raven, J. C. 1958. Standard progressive matrices: Sets A, B, C, D and E. H. K. Lewis and Co., London, 60 pp., illus.

Sadacca, Robert, Martinek, Harold, and Schwartz, Alfred I. 1962. Image interpretation task—status report, U. S. Army Personnel Research Office, Washington, D. C. *Tech. Research Report 1129*, 25 pp., illus.

Sadacca, Robert. 1963. Human factors in image interpretation. *Photogrammetric Engineering*, 29(6): 978–988.

Sims, W. G. and Hall, Norman. 1956. The testing of candidates for training as air photo interpreters. Commonwealth of Australia, Forestry and Timber Bureau, Canberra, A.C.T. 12 pp., illus.

Steel, Robert G. D. and Torrie, James H. 1960. *Principles and Procedures of Statistics*. McGraw-Hill Book Co., Inc., New York, N. Y. 481 pp., illus.