*Quantitative Evaluation of Photo Interpretation Keys**

R. DE LANCIE, W. W. STEEN, R. E. PIPPIN, A. SHAPIRO
Broadview Research and Development,
1127 Chula Vista Ave., Burlingame, California

ABSTRACT: *The procedures and findings of a year's study devoted to quantitatively evaluating certain types of P.I. keys are described. The main purpose of the study was to determine the relative effectiveness of several types of keys, particularly Dichotomous and Selective Keys. The approach was to administer a series of tests in which participants attempted to identify a number of objects, using the various types of reference material provided them. All pertinent combinations of test results were then subjected to statistical analysis. The most important conclusion to emerge is that there is no significant difference between types of keys as long as within each key the material is reasonably well organized.*

INTRODUCTION

SINCE World War II, when aerial photographs became a major source of intelligence, there have been produced a number of reference documents designed to assist the photo interpreter in extracting information from photographs. These reference documents, or photo interpretation keys as they are commonly known, have generated much discussion and even a considerable amount of controversy. Many questions have been raised as to the effectiveness of one type of key versus another, the effectiveness of keys on certain subjects, and even the value of all keys in general.

As a step toward answering some of these questions, the authors under a contract with Rome Air Development Center, U. S. Air Force, undertook a project to evaluate quantitatively certain types of photo interpretation keys. This paper sets forth the purpose, procedures, and results of a portion of that project. As it is not possible within the limits of this paper to mention all the details of a year's study, only the highlights are brought out.[1]

W. W. STEEN

PURPOSES

The major purposes of this project were: (1) to compare the effectiveness, or value, of the different ways of organizing reference material (keys) to be used by photo interpreters;

[1] A detailed final report of the entire project was submitted to Rome Air Development Center.

* Presented at 23rd Annual Meeting of the Society, Hotel Shoreham, Washington, D. C., March 5, 1957. This paper is a part of the Photo Interpretation Symposium.

(2) to determine the extent to which the value of a key type depends upon the amount of photo interpretation experience of the user; and

(3) to determine the extent to which the value of a key type depends upon the nature of the subject material.

Secondary purposes were:

(1) to determine the extent to which the value of a key type is affected by repeated use; and

(2) to determine the extent to which the value of a key type is affected by the number of objects included in the key.

Method and Background

The method used was to administer a series of tests in which participants attempted to identify a number of objects, using the various types of reference material provided them. Great care was taken to insure that all material used was technically correct and that the entire testing program was psychologically sound. Prior to the final testing program numerous preliminary tests were run in order to eliminate as many bugs as possible and to increase the effectiveness of the program. Careful test controls were designed, and uniformity was maintained throughout all the testing. All test participants, prior to their selection, were given the Moessner Floating Circles Test to insure that all were able to see stereoscopically.

In the main portion of the final testing program, ninety test participants were used, thirty from each of three different levels of experience. These three groups were:

(1) high school seniors with no P.I. experience;

(2) P.I. trainees with no working experience, represented by Air Force Officers who had just completed or were nearing completion of the five month Air Force photo interpretation course at Sheppard Air Force Base, Texas; and

(3) experienced photo interpreters with a minimum of two-and-one-half years working experience in addition to training but with little or no experience on the specific subjects tested, represented by officers and civilians engaged in some aspect of photo interpretation for one of the three services.

All participants were tested on the following three subjects: (1) Native Vegetation of the Tropical Pacific, representing

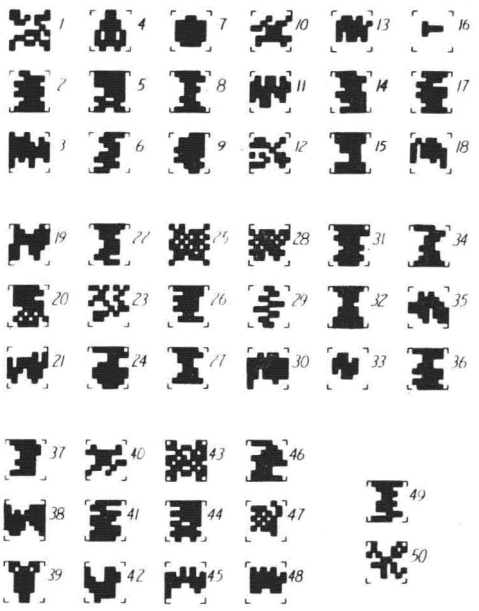

Fig. 1. Examples of abstract objects used in the main testing program.

natural features; (2) Naval Vessels, representing man-made features; and (3) Abstract Objects (shown in Figure 1). These Abstract Objects were designed especially for this study and represented the opposite end of the spectrum from natural features. The process by which these Abstract Objects are identified is similar to that employed in the identification of man-made objects on aerial photos. But the use of Abstract Objects eliminated the possibility that the test participants had any prior knowledge of the subject, as well as eliminating the influence of familiarity with stereoscopes. (Statistical analysis of all test results showed that there was a high positive correlation between the scores on the Abstract Objects tests and scores on the Naval Vessels and Vegetation tests.)

It can not be assumed that the test results for Naval Vessels and Vegetation automatically apply, respectively, to all man-made objects or all natural features. The authors do assert, however, that the results can be applied to other types of man-made objects or natural features for which the problems of recognition and identification are generally similar.

For each of the three subjects three types of reference material were used: (1) a

TABLE 1

A PORTION OF THE DICHOTOMOUS KEY USED FOR NAVAL VESSELS

A. Flight deck	See B
A. No flight deck	See D
B. Flight deck has no taper aft (square end); 1 large gun tube forward of flight deck; both forward and aft aircraft elevators square	CVE Commencement Bay
B. Flight deck has slight taper aft; no large gun tube forward of flight deck; either forward or aft aircraft elevator rectangular	See C
C. 1 large gun tube aft of flight deck (only half visible on vertical photos); aft aircraft elevator square	CVE Casablanca
C. 2 large gun tubes aft of flight deck; aft aircraft elevator rectangular	CVE Bogue
D. Wide beam in relation to length; pyramidal superstructure	BB South Dakota
D. Narrow beam in relation to length	See E
E. 3 main turrets—2 forward, 1 aft	See F
E. 5, 4, 2, or no main turrets	See I

Dichotomous form of Elimination Key (see Table 1); (2) a Selective Key; and (3) a List of Names for Vegetation and Naval Vessels, and a Disorganized Selective Key for Abstract Objects.

In a Dichotomous Key, at each step the various categories of objects are divided into two groups, based on some characteristic or characteristics which should be visible on aerial photographs. Each succeeding step subdivides the remaining group until the vessel is correctly identified. In a Selective Key all the graphic and written information pertaining to one category of object, such as one type of vegetation or one class of vessel, is presented together on a single page. The pages are then arranged in some logical order to facilitate the location of any one desired. A List of Names is just that—a list of the names of all the types of vegetation or all the classes of ships which were to be identified on the test photography. The purpose of the lists was to find out how well participants could do solely on the basis of their prior knowledge. This test served as a base for measuring the gain attributable to the use of a key.

The keys used for Native Vegetation of the Tropical Pacific were adapted from a key produced by the U. S. Naval Photographic Interpretation Center. Keys for Naval Vessels and Abstract Objects were constructed specifically for this project. All keys, except as otherwise described, covered fourteen categories of objects (i.e., there were 14 categories of Naval Vessels, 14 categories of Vegetation, and 14 categories of Abstract Objects). The Dichotomous and Selective Keys designed for Abstract Objects were very similar in format to those used for Vegetation and Naval Vessels.

For testing purposes each of the three groups of 30 comprising the different experience levels was further divided into three sections of ten each. The high school seniors and the P.I. trainees were divided on the basis of I.Q., each section having the same average I.Q. The experienced P.I.'s were divided on the basis of experience, each section having approximately the same average experience. Each section of ten was tested on all three subjects (Vegetation, Naval Vessels, Abstract Objects), but was given a different type of reference material for each subject. For example, the groups which used the Dichotomous Key for the Vegetation test then used the List of Names for Naval Vessels and the Selective Key for Abstract Objects. All tests on one subject were the same, only the reference material varied.

The test procedure was relatively simple. The member of each section, except those using the List of Names, were given approximately one hour of instruction and study on the reference material. They were then given the test material and asked to identify the objects indicated, using their keys as aids. Those using the Dichotomous Key were also given the Selective Key for that subject. They were instructed to use the Dichotomous Key as the primary

means of identification, but to check their answers with the Selective Key. Those using only the Lists of Names were given no instruction and no study period (since there was nothing to study). On all the tests there were 50 objects to be identified and a time limit of an hour and a half. When the full time limit was not used, which was the case on most of the tests, the time needed to complete the tests was recorded.

Statistical Design and Analysis

Because of the quite limited utilization of modern statistical techniques in photo interpretation, the experiments described in this paper may be of some interest from a methodological standpoint, quite aside from the conclusions to be drawn regarding keys. On this assumption, this section is devoted to comments on the statistical procedures. For purposes of brevity, only what has been named the main experiment will be discussed.

Perhaps of paramount importance from the standpoint of methodology is that no experiment (or series of experiments) should be undertaken without not only a precise statement of the purpose of the experiment, including the factors to be studied, but also a precise plan for conducting the experiment and analyzing the results. It is, unfortunately, all too common for statistical aspects to be ignored until after the experiment has been completed, at which time it is discovered that it is difficult or impossible to analyze the data because of the design employed.

Several questions should be asked and answered before selecting a design. Among them are: (1) What factors are under study? (2) What possibly influential factors not under study must be minimized, controlled, or balanced out? (3) What degree of confidence in the results is desired? (4) What experimental resources, including money, are available? And, (5) what kind of analysis is contemplated? In the case of the main experiment answers were as follows:

(1) The factors under study were types of keys, experience of the users, and the subject matter of the keys as they influence the accuracy and speed of interpretation.
(2) The factors to be minimized, controlled, or balanced out were skill within each experience level, interest, the effects of learning, and the environment and instructions under which the tests were given.
(3) It was desired that the probability of detection of differences, as large as 5% or larger between average scores, be at least .95. This was to be accomplished at the .05 level of significance (i.e., the probability for concluding that a difference between average scores existed when in fact none was present was to be .05).
(4) The resources were a fixed amount of money and a relatively limited supply of potential testees who met the requirements of stero vision, experience, intelligence, and availability.
(5) The type of analysis contemplated was a so-called "Analysis of Variance."

Ultimately it was decided to treat the main experiment as three separate experiments, one on each subject keyed, each consisting of a two-way layout. This decision was reached because it was concluded that the scores on one subject were essentially not comparable with those on another since there was no absolute standard of difficulty to go by. Within each two-way layout each person at each experience level was exposed once to each key type as described earlier. In this way the influence of the factors of interest was allowed full play while at the same time the influence of factors not under study was minimized. Thus the influences of variations in individual skill were balanced out by equalizing the groups at each experience level with respect to either I.Q. or experience; variations in interest and environment were reduced through appropriate selection of testees and by standardization of instructions and other aspects of the test environment; and finally the effects of learning were eliminated by variations in the order in which individuals were tested on the various key types.

As mentioned above, the analysis of variance was employed to draw inferences from the data obtained. In this method the assumption is made that an individual's score for a particular subject can be expressed as a sum, namely the grand mean for that subject (estimated to be the average of the scores of all 90 test participants) plus an amount reflecting the effect of experience level, plus an amount reflecting effect of the type of key, plus an

amount reflecting effect of the interaction between experience and key type, plus an amount reflecting individual variation. Except for the grand mean each of these amounts can be positive or negative; in fact, by definition of the average, there must be the same amount of negative as positive deviation from it. The great value of the analysis of variance approach is that it permits one to estimate the magnitude of these individual deviations from the average and thus to infer the influence of the factors under consideration. This is precisely what was done with the data collected and is the basis for the conclusions given in this paper.

Results and Conclusions

Table 2 contains a summary of average scores for all tests given in the main portion of the final testing program just described. Individual scores were obtained by awarding two percentage points for each correct identification. No credit was given for any incorrect identifications or for identifications not attempted. Practically all participants, except those who used the Disorganized Selective Key for Abstract Objects, attempted all fifty identifications.

All pertinent combinations of scores were subjected to statistical analysis. The results of the statistical analysis were then studied. The following conclusions resulted, subject to the qualifications stated earlier.

conclusion no. 1

A Dichotomous Key apparently has no value over a Selective Key, regardless of the experience level of the user or the nature of the subject keyed, at least when there are only 14 categories of objects to be identified. All groups, on all subjects, scored almost equally well with either type of key and took about the same amount of time to complete the tests.

A point sometimes made by proponents of Dichotomous Keys is that such keys, by systematically guiding the interpreter into a one-feature-at-a-time examination of the photographs, increase the accuracy of the resulting identifications. No evidence was found that this was the case.

conclusion no. 2

It is important that keys be logically organized if they are to be of maximum *initial* benefit. This was demonstrated in the Abstract Objects test, in which those who used the Disorganized Selective Key scored considerably lower and took longer to finish than those who used the Dichotomous and Selective Keys. The Disorganized Selective Key contains all the information in the Selective Key, but as

Table 2

Summary of Average Scores for the Main Testing Program

Vegetation			
	Dichotomous	Selective	Minimum Information
High school seniors	63.6	63.8	28.8
P.I. trainees	71.8	71.2	24.8
Experienced P.I.'s	77.0	78.6	34.8
Naval Vessels			
	Dichotomous	Selective	Minimum Information
High school seniors	59.0	60.4	17.8
P.I. trainees	65.6	66.0	23.2
Experienced P.I.'s	61.0	68.8	23.6
Abstract Objects			
	Dichotomous	Selective	Disorganized Selective
High school seniors	55.0	55.4	33.6
P.I. trainees	58.2	51.0	35.8
Experienced P.I.'s	56.8	62.2	32.2

the name implies, it is presented in a very unsystematic manner.

CONCLUSION NO. 3

The value of both the Dichotomous and Selective Keys was affected by the experience of the user only on the subject of Vegetation, representing natural features. Experienced P.I.'s scored significantly higher on the Vegetation tests than did the high school seniors, with the P.I. trainees intermediate. On both the Naval Vessels and Abstract Objects tests there is no significant difference between the scores of all three groups.

A major implication of this conclusion is that otherwise qualified but relatively inexperienced photo interpreters might be trained to do as well in the identification of man-made objects as experienced interpreters who have not been trained in that particular type of man-made object. In the identification of natural features such as vegetation, however, general experience and training in this type of work would appear to be valuable. In other words, training in the identification of, say, Vegetation of the Eastern United States might well enable a P.I. to do a better job of identifying Tropical Vegetation than a person who has had no training in the identification of any natural features.

In any event, though, it appears that by using well-designed reference material, even totally inexperienced personnel could be brought quickly to a state of fair proficiency in the identification of features from aerial photographs. The high school seniors had absolutely no P.I. experience, other than about an hour of explanation, practice and study, before taking their first test. Their performance on the tests was close enough to that of the more experienced groups to suggest that this lack of experience was no great handicap.

CONCLUSIONS BASED ON ADDITIONAL TESTS

In addition to the main portion of the final testing program previously described, several additional experiments were designed. For the first of these the Selective Key on Native Vegetation of the Tropical Pacific was carefully rewritten and reorganized into an Essay Key containing the same information. Using this Essay Key, a separate group of ten high school seniors took the same test taken by those in the main testing program. Their average score was 64.3, which is very close to the scores of those high school seniors using the Dichotomous and Selective Keys. Thus it would seem that a carefully organized Essay Key would be of the same value as the Dichotomous and Selective Keys, at least for natural features with a limited number of categories.

TABLE 3

AVERAGE SCORES FOR THE ABSTRACT OBJECTS REPETITION TESTS

	Dichotomous	Selective	Disorganized Selective
Test A	55.0	55.4	33.6
Test B	59.6	56.4	43.0
Test C	65.6	57.6	56.2

In the second additional experiment the Abstract Objects tests were repeated twice by the high school seniors participating in the main testing program. Each section used the same type of key each time. The results are shown in Table 3. Users of the Dichotomous and Selective Keys evidently obtained as much value as they could from these keys on the first test, and did not improve their scores significantly on subsequent tests. (The apparent small gain for the group using the Dichotomous Key is made insignificant by the extreme variation in their scores.) The groups which used the Disorganized Selective Key, on the other hand, gave a considerably poorer performance on the first test, but improved with practice. On the third test, Test C, this group scored about as well as the others; however, they did take significantly longer to complete the test. Thus the results of this series of tests would seem to indicate that with continued use even a poorly organized key might be about equal in value to the better organized keys, assuming the same information content.

The third additional experiment was designed to determine whether increasing the number of categories of objects keyed would affect the value of one key type relative to another. For this purpose, 28-Category keys of all three types (Dichotomous, Selective, and Disorganized Selective) were constructed for Abstract Objects. Using these keys, a test was then administered to the high school seniors participating in the main testing program. As

was the case with the 14-Category keys on Abstract Objects used in the main testing program, the scores of the sections using the Dichotomous and Selective Keys are nearly equal, while the scores of the sections using the Disorganized Selective Key are significantly lower. It should be noted in interpreting these results, however, that participants had practiced on very similar material (the 14-Category Abstract Objects tests) earlier, so that relatively small differences would not be likely to show up. Subject to this qualification, it would appear likely that even when working with as many as 28 categories of objects to be identified, there is no advantage in a Dichotomous over a Selective Key.

Summary of Conclusions

To summarize all of the foregoing conclusions, again subject to the qualifications stated earlier:

(1) Dichotomous and Selective Keys are of approximately equal value, regardless of the subject or the experience of the user, and it is very likely that other types of properly designed keys would be of similar value.

(2) Photo interpretation keys should be organized in a logical manner if they are to be of maximum initial benefit; however, with continued use even a poorly organized key might become about as effective.

(3) Dichotomous and Selective Keys on man-made objects can be used with about equal effectiveness by persons at all levels of experience. Keys on natural features, however, appear to increase in value with the experience of the user.

(4) It appears that it is unnecessary to expend great efforts to present information to P.I.'s in any particular key form or in a very highly refined and organized manner. Any reasonably well organized presentation appears to suffice.

In addition to the conclusions based on an analysis of test results, we feel that this project justified another important conclusion: that many of the controversial issues surrounding photo interpretation programs are capable of resolution by the experimental method. This study certainly does not answer all the important questions, nor was it designed to. The results and conclusions are only valid within the limits of the areas investigated. Even within the areas investigated it is possible that keys on much different subjects, using substantially different testing conditions and procedures, etc., might produce results which would lead to expansion or qualification of the conclusions stated here. But it seems clear that, within the imposed limitations of the experimental design, this project has produced valid and useful results. In that sense, it may represent an important forward step in the scientific investigation of questions concerning the interpretation of aerial photographs.

Acknowledgements

The authors acknowledge the assistance and cooperation of the Reconnaissance Branch of the Directorate of Intelligence, Headquarters, U. S. Air Force, and also of Dr. Harold Young, consultant to the Assistant for Operations Analysis, Headquarters, U. S. Air Force. Dr. Young has long been deeply interested in problems concerning the quantitative evaluation of all aspects of photo interpretation, and it was largely through his continued interest that this project was conceived.

The Origin and Distribution of United States Soils

A fine color map showing the soil texture and bedrock relationships of soils in the United States and originally designed for photo interpretation work, was published by C.A.A. in 1946. It was prepared under the direction of D. J. Belcher, now head of the firm of Donald J. Belcher and Associates, Incorporated.

Out of print for the past five years, copies of the map (size 30″×50″) may be obtained at $2.00 per copy, including postage prepaid, by addressing D. J. Belcher & Associates, Inc., 130 Forest Home Drive, Ithaca, N. Y. Arrangements have been made with CAA for reprinting and distribution, at cost. This map has remained in demand as the uses of aerial photography have increased. Its original issue was in connection with the CAA publication, *The Origin, Distribution and Airphoto Identification of U. S. Soils*, of which Mr. Belcher was the principal author.