SRI International

Final Report
Covering the Period October 1983 to October 1984

May 1985

AN AUTOMATED RV EVALUATION PROCEDURE (U)

By:

[redacted]

SRI Project 7408-12

[redacted]

Copy 1 of 1

~~Copy No. 8~~

This document consists of 28 pages.

~~SRI/GF-0275~~

SRI International
333 Ravenswood Avenue • Menlo Park, California 94025 • USA
(415) 326-6200 • Cable: SRI INTL MPK • TWX: 910-373-2046

I OBJECTIVE (U)

(U) The objective of this task was to improve and automate the remote viewing (RV)* evaluation procedures.

*(U) RV (remote viewing) is the acquisition and description, by mental means, of information blocked from ordinary perception by distance or shielding.

CONTENTS

ILLUSTRATIONS

TABLES

II SUMMARY OF RESULTS (U)

We have modified a computer-automated remote viewing analysis procedure, first developed at Princeton University, to be more responsive to the needs of the community. Our procedure is based upon defining the information content in both a RV response and its associated target as the presence or absence of a series of items (called descriptors). Various mathematical comparisons can be made between responses and targets. By defining RV accuracy as the fractional part of the target information that was correctly perceived, and defining RV reliability as the fractional part of the response that was correct, we are able to construct an RV "figure of merit " as the product of the two. The RV figure of merit is a sensitive, target-pool-independent assessment of the quality of a single, remote-viewing response.

(U) We have developed a technique to assess an analysts' RV judging ability by using a standardized test case of a series of remote viewings. Judging consistancy in a training environment is the most important factor in assessment ability. Thus, it is a requirement that the same analyst assess the information content in both the response and the target. In a training environment, an analyst would first determine the information content in all of the targets in the target pool before assessing the information content in any RV response. All of the RV assessments are done without knowledge of the particular matching target.

We have suggested ways in which *a priori* probabilities, on a descriptor-by-descriptor basis, can be used as RV assessments in the absence of any knowledge of the site. This technique requires the building of track records for each item on a viewer-by-viewer basis. As the track records begin to stabilize, we will be able to integrate the analysis techniques described in this report.

III BACKGROUND (U)

(U) Since publication of the results from the initial remote viewing effort at SRI International,[1*] two basic questions remained about the evaluation of RV data:

- What is the definition of the target site?
- What is the definition of the RV response?

For example, consider a typical IEEE-style, outbound-experimenter remote-viewing trial. After an experimenter travels to a randomly chosen location at a prearranged time, a remote viewer's (RVer) task is to describe that location. In trying to assess the quality of the RV descriptions (e.g., in a series of trials), an analyst must go to each of the sites and attempt to match responses to them. For example, while standing at a site, the analyst must decide not only the bounds of the site, but must also determine the site details that should be included in his/her analysis. While standing in the middle of the Golden Gate Bridge, should the analyst consider the buildings of downtown San Francisco, which are clearly and prominantly visible, as part of the Golden Gate Bridge target? Similarly, the RV response to the Golden Gate Bridge target might be 15 pages of dream-like free associations. A reasonable description of the bridge may be contained in the response; however, it might be obfiscated by a large amount of unrelated material. How should an analyst approach this problem?

(U) The first attempt at quantitatively defining an RV response involved reducing the raw transcript to a series of declarative statements called concepts.[2] Initially, it was determined that a coherent concept should not be reduced to its component parts. For example, a small red VW car should be considered as a single concept rather than four separate concepts: small, red, VW, and car. Once a transcript had been "conceptualized," that list of concepts constituted, by definition, the RV response. The analyst, then, rated the concept lists against the sites. Although this represented a major advance over previous methods, no attempt was made to define the target site.

During the FY'82 program, we developed a procedure to define both the target and the response material.[3] We learned that before a site could be quantified, a goal for the

*(U) References are listed in the order of appearance at the end of this report.

overall remote viewing must be clearly defined. If the goal is simply to demonstrate the existence of the RV phenomena, then anything that is perceived at the site is important. But if the goal is to gain information, then specific items at the site are important, while others remain insignificant. For example, consider an office as a hypothetical target. Let us assume that a single computer in the office is of specific interest. Suppose an RVer gives an accurate description of the shape of the office, provides the serial number of the typewriter, and gives a complete description of the owner of the office. While this kind of a response might provide excellent evidence for remote viewing, the target of interest (the computer) is completely missed. What is needed is a specific technique to allow assessments that are mission oriented.

The procedure developed during FY'82 was a first attempt at solving the mission orientation problem. In this technique, the transcript is conceptualized as described above, and a similar process is applied to the sites. A target site is conceptualized as a set of target elements, which are to be considered "mission independent." In the office example above, target elements might be: desk, safe, window, telephone, computer, and chair. A second layer of conceptualization is then applied, which is "mission specific." Each target element is assigned a number between 1 and 5 corresponding to the mission's relevance. Again, in the office example, the computer would be assigned a relevance factor of 5 (most relevant), while all other target elements would be assigned a factor of 1 (least relevant). The target elements and their relevance factors constitute the site definition and mission orientation. The final report for the FY'82 task[3] described in detail how a mission specific assessment was made. Although the procedure proved to be quite sensitive, it was nonetheless cumbersome and difficult to apply.

This report describes a major advance over the FY'82 technique. The original idea, which involves computer-automated scoring of RV data, was developed at the Anomalies Laboratory of Princeton University.[4] We have significantly extended and modified the Princeton technique and have developed procedures that can be used in actual applications.

IV METHOD OF APPROACH (U)

The overall method of approach was to begin with the Princeton group's known evaluation procedure, then determine what would be appropriate for our environment. The next step was to expand the analysis concept to be more responsive to requirements, and to integrate the entire procedure with our on-line data bases.

A. (U) The Princeton Evaluation Procedure (PEP)

(U) In general, the Princeton Evaluation Procedure (PEP) is based on comparing *a priori*, quantitatively-defined target information with similarly quantitatively-defined response information. (A complete description of this procedure can be found in Reference 4.) The procedure was developed for use as a research tool in the university environment, where complete knowledge of the target sites could be obtained. Once the target and response information was defined, the PEP applied various methods of mathematical comparisons to arrive at a meaningful assessment score.

1. (U) Target Information

(U) The definition of a particular target site (usually outdoor sites in and around Princeton, New Jersey) was contained in the yes/no answers to a set of questions called descriptors. These descriptors were designed in such a way as to characterize the typical Princeton target. By definition, the only target information to be considered in the analysis was completely contained in the yes/no answers of the descriptor questions for that site. For example, one descriptor from their list, "Are any animals, birds, fish, major insects, or figures of these *significant* in the scene?" defines the animal content of the site. The question would be answered "yes" for a zoo and a pet store target, but answered "no" for a typical campus building target. Similarly a set (30 for the PEP) of yes/no responses constitutes the target information.

2. (U) Response Definition

(U) The descriptor list for the target sites is used as a definition of the response as well. For a given RV session, an analyst (blind to the target site) attempts to answer the 30 questions based entirely on the single RV response. Using the same example above, an analyst would have to decide if a particular verbal passage or a quick sketch could be interpreted as animals or not. For some responses this might be an easy task, "I get a picture of a purple cow." Most responses, however, require a judgement, "I hear a funny sound and there may be an odd smell in the air." Nonetheless, the yes/no answers to the 30 questions constitute the only response information that will be used in the analysis.

3. (U) Analysis

(U) For a given response/target combination, the information is strictly contained in the yes/no answers to the descriptors. A binary number (30 bits long for PEP) is constructed for the target and the response descriptor questions respectively. A yes answer is considered a binary "1" while a no answer is considered a binary "0." The resulting two 30-bit binary numbers can then be compared by a variety of mathematical techniques to form a score for that specific RV session. For a series of RV sessions, a quantitative assessment is made by comparing a given response (matched to its corresponding target site) against the scores computed by matching the response to all other targets used in the series. This procedure has the added advantage of a built-in, within-group control. In other words, this assessment determines the uniqueness of the target/response match compared with all other possible matches for the series.

B. (U) Problems with the PEP

There are a number of problems with the PEP when the conditions under which the PEP was developed are no longer valid. Because we are trying to develop an RV analysis procedure that is useful both in the RV training environment as well as in applications, we have identified four basic problems with using the PEP for our purposes:

- The bit descriptors were not appropriate for our training environment.
- The PEP was not interfaced to a standard data base management system (DBMS).

- The cross-target scoring procedure was not sensitive to requirements.
- Any cross-target scoring procedure is inappropriate for a training environment.

(U) As stated above, the PEP descriptors were optimized for natural outdoor sites in the Princeton area. Because we planned to use different target material, the PEP descriptor list was completely inadequate. Having obtained the computer codes used at Princeton, we noticed that the PEP required a special on-line, within-code data base. We felt this was an inefficient way to procede because we already had most of our data in a commercial DBMS, Ingres.[5]

One of the principal problems of RV used as an adjunct to conventional collection techniques is that RVers tend to add information, sometimes called analytical overlay (AOL), to the response. If training techniques are to be developed that are sensitive to requirements, they must attempt to inhibit AOL. Specifically, any training analysis procedure must be particularly sensitive to the addition of extraneous information. The PEP was completely insensitive to this requirement.

(U) We also observed that for the purposes of training, any scoring procedure that cross compares a training response against all targets in the target pool, might penalize excellent RV simply because of the lack of target pool orthogonality (i.e., how different one target is to the next). For example, consider a typical *National Geographic* Magazine photograph of a flat desert showing few features. A very good description of this site will also match many other similar sites in the target pool. Thus, a comparison of the actual match with others in the pool will tend to reduce the score for reasons other than the quality of the particular RV response.

(U) We, therefore, felt obligated to modify the PEP in such a way to address the above criticisms.

C. (U) The SRI Evaluation Procedure (SEP)

The SRI Evaluation Procedure (SEP) was developed to address not only various RV training programs, but also the potential application of the SEP to problems. Thus, it was recognized that the SEP must contain cross comparison analytical procedures that

were sensitive to AOL, and at the same time, provide a meaningful assessment of RV responses that were independent of other targets in the pool.

1. (U) Target Information

(U) As in the PEP, the SRI Evaluation Procedure quantifies the target material into binary numbers corresponding to yes/no answers to a set of descriptors. Before any of the training programs had begun, a descriptor list was developed on the basis of the target material (*National Geographic* Magazine photographs), and on the responses that might be expected for novice RVers. Table 1 shows the 20 questions (descriptors) that were used for the Alternate Training Task.[6] This descriptor list, while applicable to a novice RV training environment, is not appropriate for either advanced training or [redacted] applications. The questions are strongly oriented toward outdoor gestalts typical of *National Geographic* Magazine material. Each descriptor list must be tailored to the application requirements. The horizontal lines separating the descriptors in groups of three are an aid in translating binary numbers (derived from the yes/no answers to the questions) into an octal shorthand notation.

(U) To illustrate exactly how a target might be coded into an octal number, let's consider a photograph of San Francisco on a clear day showing the bay, the central city sky-scrapers, and the centrally-located hill (Twin Peaks). Referring to Table 1 Bit Numbers 1, 6, 8, 9, 12, 13, 16 and 17 would all be answered "yes" and thus would be assigned a binary "1." The remaining questions would all be answered "no" and thus be assigned a binary "0." Starting with Bit Number 1 on the left, the binary number that defines the information for this target is 10000101100110011000. This representation, while convenient for computers, is difficult for humans; therefore, we convert it to the octal representation as a shorthand. Using the horizontal lines shown in Table 1 as divisions, we consider each triad of bits as a binary number ranging from 000 to 111. Table 2 shows the binary-number triad to octal conversion factors.

(U) Rewriting the above binary number with triad separations for clarity, we have 10 000 101 100 110 011 000. Using Table 2, we find that this binary number converts to 2054630_8. This octal number is the shorthand notation for all the information contained, by definition, in the San Francisco target example. All targets in the data base are coded by the same technique.

Table 1

(U) DESCRIPTOR-BIT DEFINITION

Bit No.	Descriptor
1	Is any significant part of the scene hectic, chaotic, congested, or cluttered?
2	Does a single major object or structure dominate the scene?
3	Is the central focus or predominant ambience of the scene primarily natural rather than artificial or manmade?
4	Do the effects of the weather appear to be a significant part of the scene? (e.g., as in the presence of snow or ice, evidence of erosion, etc.)
5	Is the scene predominantly colorful, characterized by a profusion of color, by a strikingly contrasting combination of colors, or by outstanding, brightly-colored objects (e.g., flowers, stained-glass windows, etc.—not normally blue sky, green grass, or usual building color)?
6	Is a mountain, hill, or cliff, or a range of mountains, hills, or cliffs a significant feature of the scene?
7	Is a volcano a significant part of the scene?
8	Are buildings or other manmade structures a significant part of the scene?
9	Is a city a significant part of the scene?
10	Is a town, village, or isolated settlement or outpost a significant feature of the scene?
11	Are ruins a significant part of the scene?
12	Is a large expanse of water—specifically an ocean, sea, gulf, lake, or bay—a significant aspect of the scene?
13	Is a land/water interface a significant part of the scene?
14	Is a river, canal, or channel a significant part of the scene?
15	Is a waterfall a significant part of the scene?
16	Is a port or harbor a significant part of the scene?
17	Is an island a significant part of the scene?
18	Is a swamp, jungle, marsh, or verdant or heavy foliage a significant part of the scene?
19	Is a flat aspect to the landscape a significant part of the scene?
20	Is a desert a significant part of the scene, or is the scene predominately dry to the point of being arid?

UNCLASSIFIED

Table 2

(U) BINARY-TO-OCTAL CONVERSION

Binary Triad	Octal Equivalent
000	0
001	1
010	2
011	3
100	4
101	5
110	6
111	7

UNCLASSIFIED

2. (U) Response Definition

(U) The descriptor list shown in Table 1 and the coding techniques described using Table 2 are prepared in exactly the same way to define each RV response. For a particular training program, however, a set of *a priori* guidelines must be defined in order to aid an analyst in interpreting the various aspects of the training procedure with regard to the descriptor list. For example, it might be correct within a given training context to advise the analyst to consider all isolated lines as a land/water interface, and set descriptor Bit Number 13 by definition. How this is done is completely dependent upon the particular training procedure in question. For an example see Alternate Training.[6]

3. (U) Analysis

The SRI evaluation procedure involves two different types of analysis:

- Target-pool-dependent analysis
- Target-pool-independent analysis (training).

(U) The first of these involves descriptor weighting that gives more or less credit in the final score in accordance with an *a priori* defined algorithm. It is within this analysis that

(U)

penalties are levied for "inventing" information that is not present at the site. The target-pool-independent analysis involves a straightforward counting system that depends upon a single target/response information comparison.

a. (U) Target-Pool-Dependent Scoring Alogrithm

(U) Consider a finite set of targets, N, each of which has been coded in accordance with Table 1. Define a weighting factor

$$W_j = \frac{1.0}{P_j} \quad ,$$

where P_j is the probabililty of occurrence of Descriptor Bit j (j = 1,20) and is given by

$$P_j = \frac{\text{the number of targets that have Bit j present}}{\text{total number of targets}} \quad .$$

The weighting factors will be large for descriptors that are not common, and small for common elements in the target pool. Table 3 shows an example of a set of probability of occurrences and weighting factors taken from the Alternate Training Task. This table was derived from a set of 112 *National Geographic* Magazine photographic targets. We see from Table 3 that volcanos (Bit 7) are the rarest item in the target pool, and are thus allotted the highest weighting factor of 9.337. While correctly remote viewing a volcano will significantly increase an RVers score, inventing one where there is none will be heavily penalized.

(U) Before we construct an assessment score for a single target/response, we must define the scoring algorithm, and determine a method by which scores can be compared. Consider a single target and RV response to that target. Suppose further that the information contained in each has been coded in accordance with methods described above. The scoring proceeds as follows. In considering a single descriptor bit, j, in an RV response, there are four possible ways to match (or not match) that bit with its corresponding bit in the target:

Table 3

(U) DESCRIPTOR-BIT WEIGHTING FACTORS FOR 112 TARGETS

Bit No.	Probability of Occurrence	Weighting Factor
1	0.4821	2.074
2	0.5089	1.965
3	0.5804	1.723
4	0.2857	3.500
5	0.1875	5.333
6	0.5893	1.697
7	0.1071	9.337
8	0.5268	1.898
9	0.2143	4.666
10	0.2768	3.613
11	0.1964	5.092
12	0.3125	3.200
13	0.5804	1.723
14	0.2768	3.613
15	0.1786	5.656
16	0.1786	5.656
17	0.1339	7.468
18	0.3482	2.872
19	0.3304	3.027
20	0.1786	5.599

UNCLASSIFIED

(U)

- The target bit and the response bit are zero
- The target bit is one; the response bit is zero
- The target bit is zero; the response bit is one
- The target bit is one; the response bit is one.

(U)

While there are a number of ways to proceed (the PEP considers them all), we will confine our discussion to that particular method of comparison that met the requirements stated above. Because it is difficult to know if a descriptor bit scored as zero is the result of correct or incorrect RV, a meaningful score can only be constructed from asserted information. Thus, the SEP only considers the case in which there is an assertive response (i.e., the RVer positively states that a particular descriptor is present). Table 4 shows the contribution to the assessment score for all four cases (single-bit comparison) shown above.

(U) We see from Table 4 that if the RVer correctly identifies a target descriptor bit, he/she is awarded a large contribution to the score if the item is rare (i.e., the probability of occurrence is small), and not as much if the item is common. Likewise, if the RVers invent an item, they are penalized more if the item is rare. To analyze the complete response, the values shown in Table 4 are added to the score--depending upon the correctness of the bit-by-bit match.

Table 4

(U) SINGLE DESCRIPTION BIT SCORING

Bit j		Contribution to Score
Target	Response	
0	0	0
1	0	0
0	1	$-\frac{1.0}{P_j}$
1	1	$\frac{1.0}{P_j}$

UNCLASSIFIED

(U) To complete the target-dependent scoring algorithm, it is necessary to normalize the score described above in such a way that comparisons can be made from session to session. In the PEP, a number of different normalizing factors were explored, but we have chosen to use the "perfect score" as our normalization.

(U) Let T_j and R_j be the value of the target and the response bit j, respectively. The most negative score possible would result from inventing all items in the descriptor list not present in the target. Conversely, the most positive score possible would result in correctly identifying all present target descriptors. Let N^+ and N^- be the most positive and the most negative score, respectively. They are given by

$$N^+ = \sum_{j=0}^{n} \frac{T_j}{P_j}$$

and

$$N^- = - \sum_{j=0}^{n} \frac{\overline{T_j}}{P_j} \quad ,$$

where n is the number of descriptors (20 in the example), and $\overline{T_j}$ is one when T_j is zero, and is zero when T_j is one. Thus,

$$S_r = \sum_{j=0}^{n} \left(\frac{T_j \times R_j}{P_j} - \frac{\overline{T_j} \times R_j}{P_j} \right) \quad .$$

(U)

For the normalize score, S, to be in the range from −1, to 1,

$$S = \frac{2}{N^{+} - N^{-}} (S_r - N^{-}) - 1 .$$

(U) To convert the normalized score for each RV session to a meaningful statistic, all sessions in a series are scored against all targets in the pool except for the matching target. Thus, for M RV sessions and a target pool of N targets, there would be (N × M) − M such cross matches. Figure 1 shows a sample distribution of scores for 4995 cross matches. The solid points are the data and the smooth curve is the best fit gaussian to the data.

(U) Having completed the cross matches and constructed the best fit gaussian, statistical Z scores are calculated from the RV session scores by

$$Z = \frac{S - \mu}{\sigma} ,$$

where μ and σ are the mean and the standard deviation of the cross-match best-fit gaussian, respectively. The Z score for each session is a measure of the uniqueness of the target/response match compared with the remainder of the target pool, and it represents the final output of the target-pool dependent scoring algorithm.

b. (U) Target-Pool-Independent Scoring Algorithm

(U) The target-pool-independent scoring algorithm makes an assessment of the accuracy and reliability of a single RV response matched only against the target material used in the session. As in the case of the target-pool-dependent algorithm, the target and response materials are defined as the yes/no answers to a descriptor list (similar to that shown in Table 1). Once the session material is coded into binary, we define session reliability and accuracy as follows:

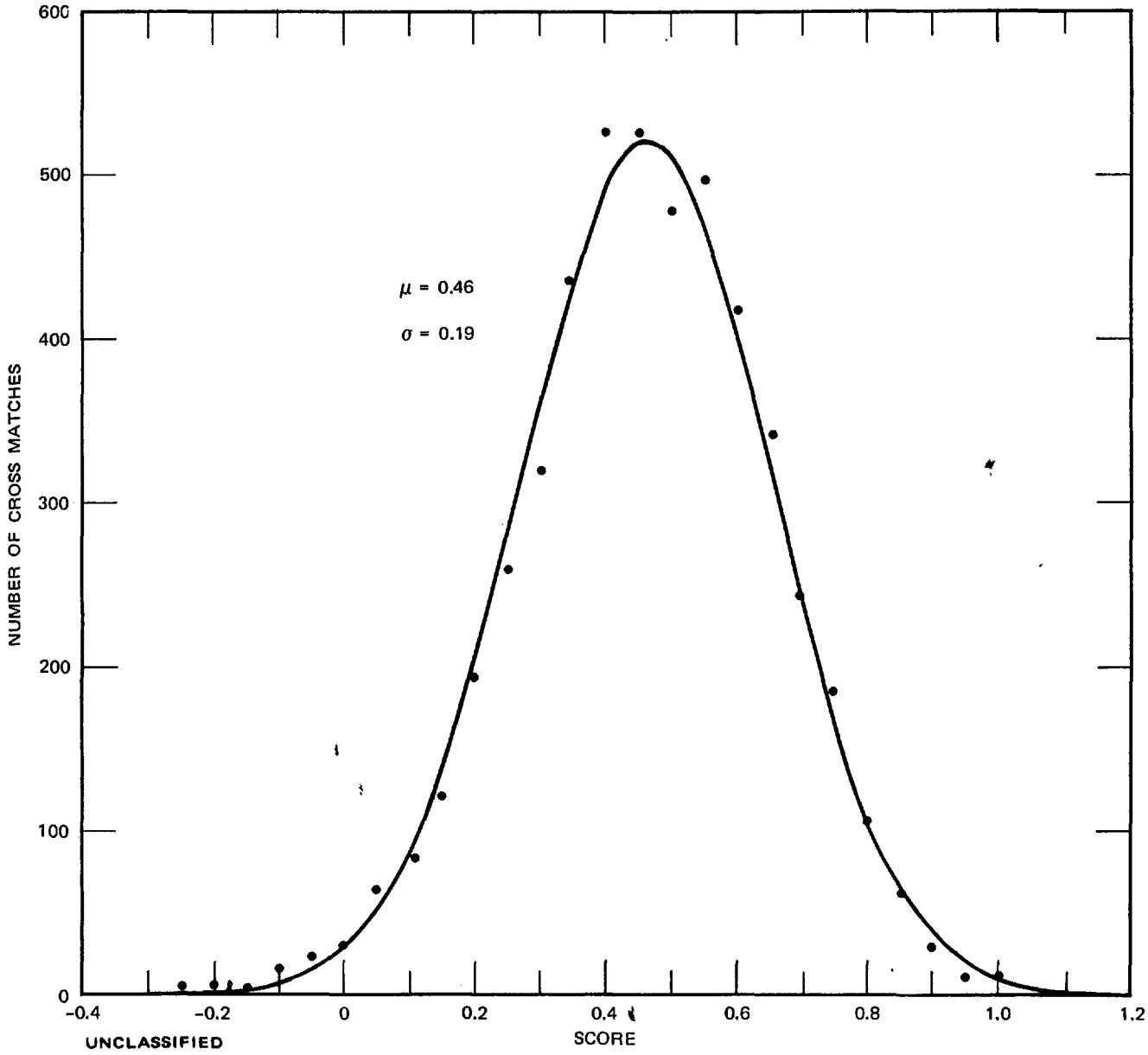

FIGURE 1 SCORE DISTRIBUTION FOR 4995 CROSS MATCHES

(U)

$$\text{Accuracy} = \frac{\text{number of correct response bits}}{\text{number of target bits = 1}}$$

$$\text{Reliability} = \frac{\text{number of correct response bits}}{\text{number of response bits = 1}}$$

In other words, the accuracy is the fraction of the target material that was correctly perceived, and the reliability is the fraction of the response that was correctly perceived.

(U) Neither of these measures by themselves is sufficient for an RV assessment. Consider the hypothetical situation in which the RVer simply reads the *Encyclopedia Britanica* as his/her response. It is certain that the accuracy would be 1.0 simply because all possible target elements would have been mentioned, and thus would not be evidence of RV. Similarly, consider a response consisting of one correct word. The reliability would be 1.0, with little evidence of RV as well. We define the figure of merit (FM) as

$$\text{Figure of Merit} = \text{Accuracy} \times \text{Reliability}$$

The figure of merit which ranges between zero and one, provides a more accurate RV assessment. In the example above where the *Encyclopedia Britanica* is the response, the FM will be low. Although the accuracy is one, the fraction of the response that is correct (the reliability) will be very small. Likewise, in the example of a single correct word as a response, the reliability is one, but the accuracy is low.

(U) A figure of merit can be calculated for each RV session to assess the progress in an RV training environment. For a series of RV sessions, the FM may be used to assess a viewer's progress on a descriptor-by-descriptor basis as well. Table 5 shows an example of FMs calculated for 22 training sessions. The "bit number" corresponds to the descriptors shown in Table 1. The "number of responses" indicates the number of sessions (out of 22) that each descriptor was asserted; the "number of targets" indicates how many targets (also out of 22) that each descriptor was asserted. The "accuracy" and "reliability" are the fraction of correct target and response material on an individual descriptor basis.

Table 5

(U) BIT-DEPENDENT FIGURES OF MERIT

Bit No.	Number of Responses	Number of Targets	Acc-uracy	Reli-ability	Figure of Merit
1	8	14	0.500	0.8750	0.438
2	1	10	0.000	0.000	0.000
3	9	13	0.539	0.778	0.419
4	3	6	0.167	0.333	0.056
5	0	6	0.000	0.000	0.000
6	13	12	0.500	0.462	0.231
7	1	2	0.000	0.000	0.000
8	14	12	0.750	0.643	0.482
9	3	4	0.750	1.000	0.750
10	0	8	0.000	0.000	0.000
11	0	2	0.000	0.000	0.000
12	2	7	0.143	0.500	0.071
13	17	15	0.733	0.647	0.475
14	6	8	0.250	0.333	0.083
15	5	6	0.000	0.000	0.000
16	1	6	0.000	0.000	0.000
17	1	2	0.000	0.000	0.000
18	3	9	0.111	0.333	0.037
19	9	8	0.500	0.444	0.222
20	0	2	0.000	0.000	0.000

UNCLASSIFIED

(U)

Finally, the "FM" is the figure of merit for each bit. For example, Bit Number 9 (city descriptor) was in the targets 4 out of 22 times. This viewer responded with "city" 3 out of 22 times. Of the 4 times a city was present in the target, the viewer correctly identified the city 3 times (thus an accuracy of 0.75). Of the 3 times the viewer responded with city, he/she was correct all the time (thus a reliability of 1.00). Therefore, the figure of merit for the city descriptor is 0.75. From the FMs of all the bits, we see that this viewer is particularly adept at remote viewing cities. Considering a large number of remote viewings, it is possible by this technique to build "viewing signatures" or track records for each viewer. When applied in the application environment, the bit-dependent figure of merit can be used as a guideline for task-specific viewer selection.

c. (U) Absolute Figure of Merit (FM)

(U) We have obtained an estimate of the meaning of FM on an absolute basis. Suppose ten viewers have contributed 50 sessions each to a training series. Each session has a figure of merit associated with it that has been calculated by the above techniques. If we add the number of responses for all viewers for each of the descriptor bits, we can obtain an estimate as to "response/analysis" bias that may have occurred across the training session. For example, suppose, of the 500 sessions, Bit Number 1 was asserted 40 times. On the average, we can assume for this training series the probability of Bit 1 being in a response is 40/500 or 0.08. By repeating this calculation for each of the descriptor bits, we can determine the probability of occurrence for all bits under the same conditions used in RV training.

(U) To determine the absolute FM distribution, a random number generator is used to create pseudo responses that are assumed to be free of psychoenergetic functioning. Each bit in a given pseudo response is generated from the emperical "bias" described above. Once the response is generated, simple descriptor-bit logical consistency is applied to finalize the pseudo response. By this technique, 10 sets of 50 pseudo responses containing no RV information can be generated. The next step is to select, on a random basis, targets from the set that were used during an actual training period to complete the pseudo sessions. The standard target-pool-independent analysis is applied to the pseudo sessions to calculate figures of merit that have, by definition, no psychoenergetic content. The histogram of FMs is fit with a gaussian distribution to provide an estimate of the mean (μ) and standard deviation (σ)

(U)

FM for random data. Since this gaussian distribution is truncated at zero FM, we must use the following procedure to determine the p–value for a given figure of merit, f.

(1) Calculate an observed z–score, $z = (f - \mu)/\sigma$.

(2) Determine an intermediate p–value, p', in the usual way given z.

(3) Calculate a normalization z–score, $z_0 = -\mu/\sigma$.

(4) Determine a normalization p–value, p_0, as in Step (2).

(5) Calculate the correct p–value, $p = p'/p_0$.

For example, during the Alternate Training Task for FY'84, the mean and standard deviation calculated as described above was 0.132 and 0.163 respectively. Therefore, using the above procedure in reverse, we find that any FM greater than 0.417 can be considered as significantly above chance.

4. (U) Testing

(U) We used the baseline data from the FY'84 Alternate Training Task to test the PEP and the SEP scoring procedures. Three analysts were asked to apply a number of techniques to the set of 6 sessions from 6 RVers each. The procedures and analysis technology that was used are summarized in Table 6. Using the descriptor list shown in Table 1, the three analysts independently scored the target pool, which consisted of 112 *National Geographic* Magazine photographs, and the set of 36 RV responses. After the scoring was completed, the three analysts met with two experienced RV judges and reached a consensus of RV quality for all 36 responses, using the 0 to 7 point assessment scale shown in Table 7.

(U) Linear correlation coefficients were calculated (using the target–dependent Z scores as the dependent variable) for Procedures 1 through 5 correlated against Procedure 7 (0 to 7 point assessment) in Table 6. From the results of these correlations, we were able to assess the effectiveness of each of the RV evaluation procedures, then determine the relative judging ability of the three analysts.

Table 6

(U) RV EVALUATION PROCEDURES UNDER TEST

No.	Procedure	Technology
1	Concept Analysis	Target/Response Concepts (equal weights)
2	PEP--Full Scoring*	Descriptor List Analysis (computer scored)
3	PEP--Selective Scoring†	Descriptor List Analysis (computer scored)
4	SEP--Full Scoring†	Descriptor List Analysis (computer scored)
5	SEP--Selective Scoring†	Descriptor List Analysis (computer scored)
6	Post Hoc Assessment‡	0 to 7 Point Scale

* Scoring includes all response bits, asserted or not.

† Scoring includes only asserted response bits.

‡ Assessment scoring done after all others.

UNCLASSIFIED

Table 7

(U) 0 TO 7 POINT ASSESSMENT SCALE

Score	Assessment Criteria
7	Excellent correspondence, including good analytical detail (e.g., naming the site), with essentially no incorrect information.
6	Good correspondence with good analytical information (e.g., naming the function), and relatively little incorrect information.
5	Good correspondence with unambiguous, unique matchable elements, but some incorrect information.
4	Good correspondence with several matchable elements intermixed with incorrect information.
3	Mixture of correct and incorrect elements, but enough of the former to indicate viewer has made contact with the site.
2	Some correct elements, but not sufficient to suggest results beyond chance expectation.
1	Little correspondence.
0	No correspondence.

UNCLASSIFIED

V RESULTS AND DISCUSSION (U)

(U) The first, and most striking, result was the necessity for the RV response coder and the target coder to be the same individual. Correlations between all scoring methods and the 0-to-7-point assessments were calculated for all possible cross-coder combinations. Only those correlations corresponding to the case where the coder of responses and targets was the same analyst, were statistically significant. This was the expected result because an analyst might be willing to adopt a liberal scoring attitude (i.e., find most descriptors present) in both the responses and targets, whereas a second analyst might adopt a conservative scoring procedure and assign few descriptors as present. As long as a particular analyst's "bias" is consistant for the targets and responses, a good assessment of RV/ability can be made. Thus, in the results described below, no cross-coder data are considered.

(U) Table 8 shows the linear correlation coefficients (which were calculated for all procedures listed in Table 6 against the 0-to-7-point assessment scale) for each of the three analysts. Because all the correlations are statistically significant, any analyst/procedure

Table 8

(U) Z SCORES CORRELATED AGAINST THE 0-TO-7-POINT SCALE

	Princeton		SRI	
Analyst	Full	Selective	Full	Selective
374	0.462	0.410	0.566	0.523
642	0.388	0.364	0.385	0.339
802	0.530	0.433	0.503	0.453

UNCLASSIFIED

(U)

combination would provide good RV assessments. The correlation coefficients averaged over all analysts were 0.431 and 0.462 for the Princeton and the SRI procedures respectively. While this difference is not significant, there is a bias in favor of the SRI procedure. Within the SRI procedure, No. 642 was the least consistent analyst. There were no significant differences between the full and the selective scoring.

In summary, we have developed a computer-based RV analysis tool that is applicable for both the training The figure-of-merit analysis allows target-pool-independent assessment of the relative progress of RV trainees. Within a given training program absolute probabilities (against chance) can be assigned for a single training session.

By carefully creating an appropriate descriptor list, and by tracking figures of merit on a bit-by-bit basis, The figure-of-merit analysis requires that complete descriptor information of the site be known. as feedback information is available, descriptor track records (figure-of-merit analysis) can be kept over many sessions to provide accuracy and reliability data on a viewer-by-viewer basis. Thus, viewers can be selected on the basis of their *a priori* probabilites on the descriptors of interest, and *a priori* assessments of their responses can be made by using the same track record.

VI REFERENCES (U)

1. H. E. Puthoff and R. Targ, "A Perceptual Channel for Information Transfer Over Kilometer Distances: Historical Perspective and Recent Research," *Proc. IEEE,* Vol. 64, No. 3 (March 1976) UNCLASSIFIED.

2. R. Targ, H. E. Puthoff, and E. C. May, "1977 Proceedings of the International Conference of Cybernetics and Society," pp. 519–529, UNCLASSIFIED.

3. E. C. May, "A Remote Viewing Evaluation Protocol (U)," Final Report (Revised), SRI Project 4028, SRI International, Menlo Park, California (July 1983) SECRET.

4. R. G. Jahn, B. J. Dunne, and E. G. Jahn, "Analytical Judging Procedure for Remote Perception Experiments," *The Journal of Parapsychology,* Vol. 44, No. 3, pp. 207–231 (September 1980) UNCLASSIFIED.

5. E. C. May, "Data Base Management (U)," Final Report, SRI Project 7048, SRI International, Menlo Park, California (October 1984) SECRET.

6. E. C. May and B. S. Humphrey "Alternate Training Task," Final Report, SRI Project 7048, SRI International, Menlo Park, California (October 1984) SECRET.