

Reprinted from the *Journal of the American Society for Psychical Research*
Volume LVI—October 1962—Number 4

Designs for Tests with Free Response Material

W. G. ROLL

A Study from the Psychical Research Foundation, Inc.

INTRODUCTION

The tests in this outline have been developed for research with "mediums" and other seemingly gifted subjects who sometimes are averse to conventional parapsychological methods, such as card-guessing trials. The designs have been adapted from the accustomed practices of these persons and are likely to be acceptable at least to those who are interested in scientific research. Since the methods allow for the free expression of alleged psychical impressions, they permit the exploration of phenomena which cannot otherwise be approached, such as those which bear on the problem of post-mortem survival.

It is recommended that a three-step test design be adopted. First, the investigator should satisfy himself in the course of informal observations that the subject appears to have psychical ability. Second, one of the designs in this outline should be selected according to the type of phenomena produced by the subject. An inventive experimenter will be able to extend the basic designs to other situations and types of test if the need arises. For instance, the methods described in Sections I and II can be used for studies of "psychic" or "spiritual" diagnosis and cure. Blood, sputum, or other samples from patients (or their names) can be used in the place of token-objects. A short pilot experiment should then be carried out with the selected design. If this is successful, a confirmatory series should be planned and conducted. It is essential in this third phase that the investigator decide on the length and type of test beforehand and closely follow his plan.

In the conduct of each type of test described in this paper, one experiment consists of four or more tests which are evaluated together (as shown in Section IV). For convenience, such a series is referred to as a "set." In pilot experiments (the second step) one set suffices, though more may be required to reach a judgment regarding the subject's abilities. As a safe rule for the third step, or confirmatory work, about fifty individual tests should be carried out before a satisfactory statistical appraisal can be made. If each set consists of four tests, twelve sets are required; if there are five tests to a set, ten sets are needed, and so on. The number of tests in each set should remain the same in a series of sets.

Reprinted from the *Journal of the American Society for Psychical Research*
Volume LVII — January 1963 — Number 1

Errata

Mr. Walter A. Carrithers, Jr., writes that Footnote 6 on page 133 of his article, "Madame Blavatsky: 'One of the World's Great Jokers'" which appeared in the July, 1962, number of the *Journal* should conclude, "Nov., 1884, p. 71."

In "Designs for Tests with Free Response Material" by W. G. Roll in the October, 1962, number of the *Journal*, the annotation mark for "correct and particularly significant" (⊙) was left out by the printer in the following places:

Page 187, Sample 1, line 10:

"⊙ for correct and particularly significant"

Page 187, Sample 2, lines 5-6:

"a man's voice (✓) spoke about insulin-shock or -shot (⊙). The person's hand feels numb (⊙) . . ."

Page 188, 2nd main paragraph, lines 3-5:

"A ⊙ he can count as 2 or more, as he decides, only he must decide on the value he assigns it before the tests. I generally count ⊙ as 2."

Page 189, bottom line of Section II:

"No. 2 would rate a ⊙ if true"

Page 189, line 2 of Record Sheet:

"Write ⊙ for correct and particularly significant"

Page 193, line 3 from bottom:

"With ⊙ counted as 2, the score is 7."

PRINTED IN BRUGES, BELGIUM, BY THE ST CATHERINE PRESS, LTD

Though the number of tests in each set must be the same, the subjects and cooperators may be different. A set consisting of four to six tests generally can be completed in about one hour.

I. TOKEN-OBJECT TESTS

In token-object tests (popularly known as "psychometry"), the subject, by touching or looking at an object, allegedly can obtain information which refers to events associated with it or with the life of some person who has been in contact with it.

In each experimental set, one subject takes part, and there should be four or more cooperators. Each submits a token to the experimenter, this object belonging either to himself or to deceased or living relatives or associates. The cooperators also consent to annotate the statements produced by the subject.

It is important that the subject should not be able to infer anything from his sensory contact with the object which could contribute to the score. To this end, the objects must be either concealed (say in boxes or envelopes of similar shape and weight) or, if not concealed, must all be of the same kind. For instance, hair tokens from persons of the same sex and age groups can be used. Although the subject may infer from a perfume scent that a lock of hair belongs to a woman (or from a grey streak, to an elderly person), this information will not help towards a higher score since all cooperators in the test are likely to accept the related statements as applying to them. It is also important that the experimenter or others present during the test should not know which token-object is being used at any one time, since they might unwittingly direct the subject's responses. If hair tokens from persons familiar to the experimenter are used, he must arrange to have the subject work behind a screen, or in some other way insure that he (the experimenter) does not see the hair; and he must instruct the subject not to refer to its color, or the eye and skin color of the person to whom it belongs.

All tokens should be in similar containers, whether or not they are to be taken out during the test. Hair tokens should be tied with similar pieces of material. Each container should have an identification mark, and a code sheet should be kept, showing which name is associated with which container and enclosed token. The experimenter and others present at the test must not know this sheet. These conditions are easily met with if the experimenter has an assistant. If not, the experimenter may request the cooperators each to select a numbered container, out of sight of the experimenter, to place the token in it, and to write down on the code sheet which mark the container of his token has.

If concealed tokens of varying weights are used, these are placed by the experimenter in the lower halves of a set of similar boxes; the boxes are weighed and wads of cotton or similar material are placed in them in different quantities to equalize the weights and prevent rattling. One or more of the cooperators are then called in and asked to place the numbered tops on the boxes and to make a record on the code sheet of which box contains which object. This is done out of sight of the experimenter, and the code sheet is not consulted by him until after the experiment.

The tokens can be given to the subject by the experimenter (one object at a time) at the same session or they can be spaced over several.

The subject's statements are recorded by hand or recording machine. At the end of a set, the experimenter edits the statements. For instance, if locks of hair are used, any reference to the hair, eye, or skin color is deleted, since this would tell the cooperators which statements were intended for their tokens. (If the experimenter knows the cooperators and was himself present during the tests, the subject should not have made any such references, as noted earlier.) Repetitions are also deleted. Parentheses for the annotations of the cooperators are then inserted after each item of information. No firm rules can be given for this. The experimenter must be guided by his own judgment as to what constitutes separate items of information.

Finally, copies of the statements from all the tests in a set are prepared for each cooperator and sent to him for annotation, with indications as to the time of the tests. The name of the cooperator is written on top of the first page of his copy of the statements, and a space reserved for the date of annotating the transcript. The cooperators are also given instructions for annotating the records. A sample of this is included here (see Sample 1) and will explain the annotation marks. A sample of a record sheet from a set consisting of four tests also is included (see Sample 2). Remember that there would be four copies of these statements, one to each cooperator. This sample has the annotation marks of the cooperator to whom it was sent and also the identifying marks and scores (in italics) filled in by the experimenter, as explained below, after the record sheet was returned to him.

Sample 1

Instructions to Cooperators in Token-Object Tests

Here is a copy of the statements produced by the subjects during the sessions making up this experimental series. As a rule, the statements appearing together are intended to apply to the

same token. However, it is possible that some statements and whole paragraphs may refer to another, so you are asked to examine carefully each statement, determining whether or not it is applicable to the circumstances of the person whose token you supplied, even if the statement occurs in the midst of sentences which do not apply.

The pairs of parentheses following the statements of the sensitive are to be used for your annotations. The following marks should be used:

✓ for correct and particularly significant

✓ for correct

X for incorrect

? for doubtful or unverifiable

Each annotation should refer to the information contained in the words immediately preceding the parentheses in question.

Please do not add or remove any parentheses, but feel free to give your comments on the statements and the way they are subdivided on an accompanying sheet.

Try to read all the statements as if they were meant for you, judging them correct, incorrect, etc., accordingly.

Please make an entry in every pair of parentheses. At the end, glance over the record to make sure that every pair of parentheses has been annotated.

When finished, please write the date and time of annotating the statements in the space reserved.

Your responses to the statements will be kept confidential.

Sample 2

Statements from Four Token-Object Tests (with annotation marks and scores added)

Cooperator: **Mr. Peter Gimble**: D. Time of Annotations: **March 30, 1961, 5 p.m.** Time of Experiment: **March 9, 1961, 3:30 p.m.**

A man's voice (✓) spoke about insulin-shock or -shot (✓).

The person's hand feels numb (✓), it draws (✓), and hurts (X). The feeling goes up to here (subject indicates area just below elbow) (X). Numbness starts slow (X) and builds up (X). It's in both hands (X). I associate it with a man (✓) because of the man's voice.

Time of Experiment: **March 16, 1961, 5 p.m.**

Melancholy (✓), sad (✓) feeling. A recent death or tragedy or something very sad (?). An almost dark room (X). Almost a feeling of grief (X). Sense of loss (X). A death, or someone

going away (X). Sample belongs to a woman (X), a female personality (X).

Time of Experiment: **March 16, 1961, 8:30 p.m.**

An impression of a royal blue castle (X) and a black lacquer table (X). A woman (X) with dark but not black hair (X). Something wrong between the throat and waist (✓).

Time of Experiment: **March 27, 1961, 7 p.m.**

Person has a peeling (X), sunburnt (X) arm. Not a bad burn (X). There is a pain in the jaws (X). Person has been to a dentist (X). The walls in the dentist's office are turquoise (X), a pale shade of turquoise (X). A feeling of wanting to go to sleep (✓), on the verge of blacking out (X).

When the records have been returned by the cooperators, the experimenter adds up the score for each test. He disregards the X's and P's, and counts a ✓ as 1. A ✓ he can count as 2 or more, as he decides, only he must decide on the value he assigns it before the tests. I generally count ✓ as 2. The score of each test is written in the bottom left-hand margin after the statements on each copy. Then and only then, the experimenter determines which token-object was used in which test by consulting the code sheet. The identifying mark (A, B, C, or D, in the present sample) placed on the container of the token-object of each cooperator is copied after his name on the top of the first page of the copy of the statements sent to him. The identifying marks are inserted on the top left-hand margin of the statements from each test in a set (see Sample 2). The hope is, of course, that each cooperator will find in the test many correct statements devoted to his object, but few in the others.

II. "BILLET-READING" TESTS

This method is intended for tests with a subject working with a group of cooperators present together. It provides a scientific framework for a common "séance room" situation.

The following is an adaptation of so-called "billet-reading," but other token-objects than written sheets of paper may be used as long as these are identical in order to insure that their owners cannot be normally discovered either by the subject, the experimenter, the cooperators, or others present during the test.

One session is devoted to an experimental set of four or more tests. One subject and four or more cooperators participate. The material consists of a tape recorder (hand-written notes may be used instead), identical blank sheets of paper, opaque envelopes, and record sheets which consist of series of columns with spaces

numbered 1-100, and room for the participant's annotations (see Sample 3). If there is any question as to the opaqueness, the sheets should be enclosed in cardboard, aluminum foil, or a similar material before being placed in the envelopes.

Each of the participants writes his name, the name of a deceased person, and a question or whatever it be, on a sheet of paper. The experimenter collects these, goes into another room and inserts each sheet into an envelope, out of sight of the subject and the cooperators. It is important that he must not himself have any way of knowing which sheet is in which envelope. He should therefore use a screen or similar device when carrying out this task. If the experimenter has an assistant, this task can be delegated to him. (The assistant does not enter the experimental room or communicate with the subject or participants till after the tests.) The envelopes are then sealed and marked, e.g., with the letters of the alphabet.

After such preliminaries or introductions to the group as the subject may be accustomed to, the experiment can begin. The tape recorder is turned on (or the note-taker is ready) and the experimenter gives the subject the first envelope. The subject (or the experimenter) reads the code mark on the envelope into the tape for later identification. The subject then produces his statements pertaining to the envelope. The experimenter must always be in a position to watch that the envelopes are not being tampered with. This procedure is repeated for each envelope, after which they are returned to the experimenter.

It is then time for the cooperators to make their annotations and a record sheet (see Sample 3) is given to each. It is explained that the tape recorder will shortly be turned on for a play-back (or the notes read aloud) and that, after the first statement has been heard, an annotation mark should be placed opposite No. 1 on the record sheet. For instance, assume that the subject has been asked to describe a deceased relative of each of the participants and that the first statement is "I see a woman." Those participants whose relative was a woman would be told to enter $\checkmark$ opposite 1, the others, X. If the next statement were "She has an elaborately carved comb in her hair," No. 2 would rate a $\checkmark$, if true.

Sample 3

RECORD SHEET

Cooperator: *Peter Gimble*. Time of Test: *March 9, 1961; 3:30 p.m.*
Write $\checkmark$ for correct and particularly significant, $\checkmark$ for correct, X for incorrect, and ? for doubtful or unverifiable. Draw a line

beneath the last number of each test. If there are more than 100 items to be annotated, use another record sheet, changing 1 to 101, 2 to 102, etc.

D	1	$\checkmark$	21	X	41		61		81
	2	$\odot$	22	X	42		62		82
	3	$\odot$	23	$\checkmark$	43		63		83
	4	$\checkmark$	B 24	X	44		64		84
	5	X	25	X	45		65		85
	6	X	26	X	46		66		86
	7	X	27	X	47		67		87
	8	X	28	X	48		68		88
	9	X	29	X	49		69		89
	10	$\checkmark$	30	X	50		70		90
A	11	?	31	$\checkmark$	51		71		91
	12	X	32	X	52		72		92
	13	$\checkmark$	33		53		73		93
	14	X	34		54		74		94
	15	$\checkmark$	35		55		75		95
	16	X	36		56		76		96
	17	X	37		57		77		97
	18	X	38		58		78		98
	19	X	39		59		79		99
C	20	X	40		60		80		100

For Experimenter's Use
Count $\checkmark$ as 1, count $\odot$ as 2

Tests	D	A	C	B		
Cooperator's Test	D					
Scores	7	2	1	1		

The experimenter decides how to separate the material into items for annotation, as in other token tests (see Section I).

Once the instructions have been given, the play-back and annotation can begin. The experimenter stops the tape (or ceases reading from his notes) after each statement, to give the cooperators time for their annotations. At the end of the statements pertaining to the first envelope, they are told to draw a line beneath the last annotated number on their records, and to write the identifying letter of the next envelope by the following number, and so on to the end. The record sheets are then collected by the experimenter after he has made sure that they are filled in correctly.

The experiment has now been completed and it remains only to open the envelopes to determine who is associated with which. The letter on the envelope of each cooperator should be copied onto his record sheet. It will probably not be convenient to complete the assessment at the time of the test. However, the participants will get a good idea of the results if they are told which envelope and identification mark is theirs, and if the tape is then played back again (or the notes again read aloud). The items scored correct by each cooperator for each envelope are then added by the experimenter, using the spaces at the bottom of the record sheets.

III. ESP-PROJECTION TESTS

In ESP-projection ("traveling clairvoyance" or "out-of-the-body" experience), the subject has the experience of being in a place different from where his body is and of observing events there which, later it is found, occurred. Sometimes the objects allegedly observed include the subject's body; and occasionally it is claimed that the subject was seen by others at the place where he had the impression of projecting himself.

Reported cases of ESP-projection generally are associated with changed states of consciousness brought about by drugs, sleep, hypnosis, or self-induced dissociation. However, the tests in this guide are designed for investigating subjects who already possess psychical abilities of the types in question; the conditions which may bring about such capacities are not considered.

The experimental design described here is equally suitable for other free response ESP tests. An experiment (a set) consists of four or more tests. In each, the subject is instructed to project himself to a different location or to the same location at different times, and to report what he observes to the experimenter, who keeps a record. One or more cooperators who can observe or are familiar with the target locations assist in the experiment. They are provided with copies of the subject's statements at the end of the experiment and requested to annotate these, without knowing

which statements are meant to apply to which circumstances. The cooperators either receive a full transcript of the statements (similar to the one inserted in Section I) or the statements are played back by a tape recorder, or read aloud and annotated on a record sheet as explained in Section II.

Occasionally a subject repeats part of the experimenter's instructions in such a way that the cooperators might know or be able to guess where the subject attempted to project himself, if they were to see this part of the transcript. Such statements must be deleted or edited by the experimenter before making out the copies for the cooperators. For instance, assume that the subject has been instructed to visit a room in a hotel and he remarks that "the lobby is crowded with people." The cooperator who is associated with the hotel will at once know that the record where this sentence appears relates to the hotel. For this reason it has to be deleted. If another target location is a cabin on a ship and the subject says that "the bunk is under the porthole," this, too, will tell the cooperators which location the subject is aiming for, and the statement must be deleted (or edited to "the bed is under the window"). It is not always easy to decide whether all such clues have been removed. It is best therefore to equalize the target locations. For instance, if it is desirable to use a hotel room, all target locations in the experiment should be hotel rooms, or all should be living rooms, or all bedrooms, etc.

In a convenient adaptation of this procedure, there is only one cooperator and the results can be assessed immediately. The subject is asked to project himself to a certain room where he will observe the cooperator, and to report back on his activities. The experimental set is divided into four or more tests of, say, ten minutes each. At the end of a set, the statements are copied for annotation, or, more simply, read aloud to the cooperator who can annotate them by using the kind of record sheets shown in Sample 3. The cooperator uses one record sheet for each test period. If a set consists of four tests, all the items in all tests are annotated according to the accuracy with which they describe the cooperator's activities in each period. For instance, if during the first period he sharpened a pencil while smoking a cigarette, he annotates on the first record the subject's statements for each period according to whether they seem to refer to smoking or pencil sharpening; on the next record sheet he evaluates the items according to his activities during the second period, and so on. He must keep a record of his activities during each period as a guide for his annotations.

Before the annotation can begin, it is important that the experimenter remove any clues from the records that can inform the

cooperator which statements were meant for which periods. For instance, if the subject says, "He is no longer sitting down," the cooperator, when he later sees or hears this statement, will infer that it was *not* said during the first period, and the annotation and assessment may be biased as a result. For the same reason, the order of the tests must be randomized before being presented to the cooperator.

IV. CALCULATING THE RESULTS

The best way at present to determine the statistical significance of tests with verbal material is the Pratt-Birge method. The following description has been adapted from an article by J. G. Pratt.¹

When the experimenter has completed adding the scores for each of the tests of a set, he enters them in a table, as in the following example where four cooperators took part (see Sample 4). He sums up the numbers in the rows and columns and in the "target diagonal." He is then ready for the analysis, a standard statistical procedure, easy for anyone to complete (a table with squares and square roots will, however, be a help).

Sample 4

Cooperators	Cooperator's Annotations			
	Smith (A)	Olson (B)	Peters (C)	Gimble (D)
Smith (A)	6	0	0	1
Olson (B)	2	1	4	1
Peters (C)	0	2	4	3
Gimble (D)	2	1	1	7
	10	4	9	12
				35

The spaces in Sample 4 are made up of the number of items checked as correct by the cooperators. For instance, Gimble found 5 correct items in the statements the subject produced when this person held Gimble's token or projected himself to Gimble's location; (at the time of annotating these statements, Gimble did not know that they were intended for him). Of these, two were "correct and particularly significant" (see samples of statements and record sheet in Sections I and II). With $\checkmark$ counted as 2, the score is 7. He also found 2 correct items in Smith's record, and 1 each in Olson's and Peters'.

¹Appraising Verbal Test Material in Parapsychology, *Journal of Parapsychology*, Vol. 12, December, 1948, pp. 236-256.

The total number of items is 35. Of these, 18 were hits (the sum of the items on the diagonal). The mean number of hits expected by chance is one fourth of the number of items, that is $35/4$ or 8.75. In this experiment the deviation from chance is 18 less 8.75, that is, 9.25.

Some additional figures are needed to complete the assessment:

- (1) The square of the number of items (I) = 1225.
- (2) The sum of the squares of the individual spaces (S) = 143.
- (3) The sum of the squares of the separate row totals (R) = 315.
- (4) The sum of the squares of the separate column totals (C) = 341.
- (5) The number of tests (t) = 4.
- (6) The square of the number of tests (T) = 16.

$$\text{Variance} = \frac{I + (T \times S) - (t \times R) - (t \times C)}{T \times (t - 1)}$$

In the present experiment this works out at:

$$\frac{1225 + (16 \times 143) - (4 \times 315) - (4 \times 341)}{16 \times (4 - 1)} = \frac{889}{48} = 18.52$$

If the experimental series consists of 12 sets (48 tests), the 12 individual variances are simply added together, as are the 12 deviations. Let us say that the total deviation in the 12 sets is 36 and the total variance is 225. The critical ratio, which indicates the statistical significance of the result, is obtained by dividing the total deviation by the standard deviation. The standard deviation is easily calculated by taking the square root of the total variance ($\sqrt{225}$). The critical ratio (CR) thus works out at $36/15 = 2.40$.

Earlier in this paper I suggested that the experimenter should complete one or more exploratory sets before engaging on the full series necessary for a statistical assessment. I have not considered the question of how large the deviation in a single set should be in order to justify conclusive testing. An average deviation of 3 in the 12 sets (assuming a total variance of 225) would yield a critical ratio of 2.40. In other words, if the exploratory set described above had given a somewhat smaller deviation than the one shown (9.25), further testing would still have been indicated. If there are more items or more cooperators in an exploratory set, the deviation has to be proportionately larger to be suggestive.

A critical ratio of 2.33, which has odds of 100 to one, is usually considered statistically significant. If the experimenter is prepared for negative scores as well as positive ones (that is, if he suspects that the subject may *avoid* giving correct impressions), a critical ratio of 2.58 has to be reached to obtain the 100 to 1 level.

CONCLUDING REMARKS

This paper is part of an attempt to encourage research with so-called "sensitives," "mediums," and other gifted subjects whose abilities have been neglected in recent years, partially because they often operate in religious and occult areas, with which the parapsychologist is anxious not to appear to be associated, and partially because there have been no suitable methods of assessment. This field of research is important not only because the potential subjects often appear to possess remarkable and recurrent abilities in paranormal cognition but also because these powers may throw light on problems which have not yet been closely studied, such as the question of post-mortem existence.

A guide for tests with run of the mill subjects would be incomplete without a section on the psychological conditions favorable to success. The investigations outlined here, however, are primarily intended for persons who already are known, or suspected to possess, psi abilities. The main purpose of the tests is to allow for scientific assessment without radically changing the favorable conditions under which the results are produced.

This is not to say that anybody can get scores with these procedures, even from capable percipients. The subject's knowledge that he is being observed, the personality of the investigator, and other factors may hamper results. For this reason a three-stage approach is recommended. This offers an opportunity, first of all, for the investigator to study the subject under the percipient's normal working conditions and to ascertain whether results can be forthcoming while he is being observed. At the second stage, the appropriate test design is applied with such adjustments as seem advisable. If the subject continues to produce under these conditions, a confirmatory series is initiated. If the experiments at the second stage of the research are unsuccessful, but there were strong indications of psi in the first stage, further search for a fruitful method may be undertaken. There is wide scope for improving and adding to the designs.

**Psychical Research Foundation, Inc.
Box 6116
College Station
Durham, North Carolina**