1961

The reliability of Rorschach movement responses.

Gail Kathryn Lambert
University of Massachusetts Amherst

THE RELIABILITY OF RORSCHACH MOVEMENT RESPONSES

LAMBERT

THE RELIABILITY OF RORSCHACH MOVEMENT RESPONSES

Gail Lambert

A thesis presented in partial fulfillment of the requirements for the Master of Science Degree

Department of Psychology

University of Massachusetts, Amherst

December, 1961

Table of Contents

Introduction

Review of Rorschach Reliability

Despite numerous studies on the reliability of the Rorschach test, no satisfactory conclusions have been reached. The split-half method, used in several studies, is an unsuitable one because of the small number of "items" represented by the ten blots, and because the distinctly different natures of the blots makes it impossible that the two halves be equivalent. Vernon's results as reported by Hertz (1934) showed highest split-half reliability for number of responses to be .91, and the rest of the coefficients to vary from .33 for F+% to .74 for W%. For M, reliability was .62. Reliability was higher for records containing more than 30 responses than for those with less; it was therefore concluded that total R should be held constant (Klopfer, Ainsworth, Klopfer & Holt, 1954, Ch. 14). Cronbach (1949) has also stated that controlling for total number of responses is important. Hertz (1934) found split-half reliability coefficients that ranged from .76 for M% to .89 for W%, attributing the difference between her results and Vernon's in part to the more highly standardized conditions of administration in her study. She did not, however, attempt to hold total R constant. As a result, some of her ratios have varying denominators. According to Cronbach (1949) the Spearman-Brown formula which Hertz used cannot be used to test ratios

with varying denominators.

Thornton and Guilford (1936) tested the split-half reliability of Erlebnistypus scores. They tested two groups of students under different conditions: group I with scant instructions and inquiry; group II with fuller instructions, inquiry, and time limits. For group I, reliability for M was .919 and for sum C .938. For group II, M reliability was .768 and for sum C .655. The M/Σ C ratios, .814 for group I and .307 for group II, the authors admit are probably spurious, since computing correlations between ratios such as M1/C1, M2/C2, etc., is statistically unsound. Also, if two scores are unreliable, their ratio will be even more unreliable. It was not possible to find out what caused the drop in reliability from group I to group II.

Results from the test-retest method are subject to distortion through the influence of a memory factor. Kelley, Margulies, and Barrera (1941) tried to circumvent the memory factor by retesting patients who had just undergone electroconvulsive therapy, and who, though free from confusion, had complete amnesia for a Rorschach test administered just prior to the shock. The psychograms and the diagnostic impression gotten from them seemed largely the same to the authors. The only shifts of more than one response were found in total R, d, F%, (VIII, IX, X)% and P. Unfortunately, only twelve subjects were included in the sample, and therefore the authors did not attempt any statistical evaluation of the data.

Holzberg and Wexler (1950) did a test-retest study with 20 chronic schizophrenic patients, hypothesizing that high reliability coefficients could be obtained for the behaviorally unpredictable schizophrenics only if the Rorschach measures a stable, underlying personality organization. They found reliability coefficients to be significant at the .05 level for most of the scoring categories. Their use of such a small, homogeneous sample of subjects makes their study more an attempt to set limits of reliability than an attempt at measuring reliability per se.

Fosberg (1941) set out to test the Rorschach's vulnerability to faking as well as its test-retest reliability. He gave each of 50 subjects four administrations of the Rorschach: the first and fourth under standard conditions, the second with instructions to try to make the best possible impression, and the third with instructions to make the worst possible impression. He found correlations in the .90s for the two tests given under standard conditions, and the lowest correlations, those between the "best" and "worst" tests, were in the .80s. He concluded that the Rorschach was highly reliable and unfakable. His conclusions are highly suspect, however. He did not attempt to control for total R. As Cronbach (1949) points out, his statistical techniques are unsound. He calculated test-retest correlations for location, determinants and content separately, as well as for the test as a whole; that is, two sets of scores for one person were

correlated, pairs of values such as D1-D2 (number of large detail scores for the first and for the second administrations) were entered in the same chart. But, since each score has a relatively limited range for all people, e.g., more detail than whole responses tend to be given, the greater magnitude of D causes the two sets of scores to correlate -- a high correlation would also have been gotten if the scores had come from different subjects.

Most researchers agree with Zubin (1954) that an alternate form of the Rorschach is needed to test reliability. In 1920, Behn-Eschenburg, in collaboration with Rorschach, developed the Behn-Rorschach series. The blots are similar in construction to the Rorschach and were designed to be equivalent to it. In 1941, Zulliger prepared a manual for use with the Behn blots. Zulliger does not, however, include in his manual any quantitative data on the relationship between the two series.

Eichler (1951) found correlations of .50 for M and .51 for FM testing with the Behn and Rorschach parallel forms. He concluded that the Behn showed substantial agreement with measures obtained from the Rorschach, but that correspondence was not close enough to differentiate between individuals. However, Eichler made no adjustment for variance due to general productivity, that is for total R. Also, it is possible that some of the significant differences found between mean scores may be falsely significant; 36 significence tests

were done, which may have inflated the probability values to some extent.

Singer (1952) found correlations of .89 for [illegible] and .81 for number of [illegible]. He concluded that the correlation of scoring categories between Behn and Rorschach is high enough for group prediction, but not high enough for individual prediction. However, the study is far from a definitive one, as Singer himself admits. The scores of his first group of ten subjects were tested for significant differences using a critical ratio measure, which is a statistically unsound technique for such a small sample. The time interval between tests was not held constant. A second estimate of reliability was made with 23 cases from Zulliger's manual added to his experimental group. It was impossible to completely equate the scoring for the two groups, and discrepancies in scoring tend to lower correlations between categories. Also, Singer did not control for total number of responses.

Schwartz and Yates (1957) compared the equivalence of the Behn and the Rorschach tests under standard and stress conditions for matched groups of homogeneous subjects. The statistical analysis treated the control groups (standard conditions) and experimental groups (stress conditions) as separate, one-dimensional designs, thus allowing the computation of parallel for reliability coefficients. They found correlations of .45 for M, .53 for FM, and .72 for m. The Behn and Rorschach were found significantly different on only

two variables from among 16 comparisons that were made, the Behn eliciting more FK and FC responses than the Rorschach. None of the correlations was as high as .75, which is at best a minimally acceptable figure for individual prediction. One of the movement variables, m, does come close to this criterion, with a correlation of .72. However, as Schwartz and Kates state, the coefficients are based on only 12 subjects, and must therefore be interpreted with caution.

Epstein, Nelson and Tanofsky (1957) constructed 100 inkblots, which were assigned randomly to ten sets of ten cards each. They administered the sets of blots to 16 subjects, giving them in two sessions a week for a period of five weeks. In order to hold total R constant, three responses per card were required. They used several scoring categories of their own as well as standard Rorschach scores. They found correlations of .42 for M, .27 for FM, and .23 for m. For the results over all ten sessions, every score measured individual differences to a statistically significant degree, but the highest reliability coefficient obtained was only .56. The authors suggested that different scores and combinations of scores might have yielded more reliable results.

Movement Responses as Relating to Energy Expression

A number of studies exist which suggest that scores which reflect energy level of movement might provide more reliable results than those previously cited. Singer, Meltzoff and Goldman (1952) measured the number of movement

responses given by subjects tested before and after a period of hyperactivity (five minutes of vigorous calisthenics) and before and after a period of motor restraint (remaining "frozen" for five minutes in an uncomfortable position). No difference was found after hyperactivity, but after the motor restraint, there was a significant increase in number of M responses, and in number of FM plus m responses.

Meltzoff, Singer and Korchin (1953) found significantly more movement responses when testing after a motor inhibition situation, in which subjects were made to write a phrase as slowly as possible. They were interested primarily in M, though they used in one part of their experiment a combined FM plus m score, and in another a total movement score, without finding conclusive differences in effect of inhibition on the three types of scores. Some doubt attaches to their conclusions: in all three parts of their experiment two statistical measures were used; a correlation measure which was significant, and an analysis of covariance which was not significant, though the differences were in the predicted direction. On the other hand, it should be taken into consideration that only gross number of movement responses was measured, and that only two to four cards were given. Increased significance might have resulted if a larger number of cards had been administered, and if the movement responses had been weighted for intensity.

In general, then, it seems that the results of the above

experiments indicate that the Rorschach movement response is related to energy expression.

Recent research suggests the importance of the relative amount of activity in the movement percept. Singer and Spohn (1954) compared schizophrenics with introversive and extratensive M/ΣC scores on a motor inhibition task and rated their activity level during a fifteen minute waiting period. They found longer inhibition times and lower activity levels for the group with high number of M. Then, the responses of the high M group were divided into "active" and "static" movement. Fifteen subjects with two or more active M and ten with less than two active M constituted the two subgroups. The active M subgroup had longer inhibition times ($p = .01$) and lower activity ratings ($p < .02 > .01$), and the same was found when the total M sample was broken into active and static groups. Singer and Spohn thus drew the conclusion that vigorousness of movement was a variable deserving consideration. Further study of this variable looks even more promising in view of the fact that Singer and Spohn, using a simple "active-static" classification, found such a clear-cut difference in a homogeneous population giving only few M responses.

Zubin (1948) constructed scales for use with the Levy Movement Blots. His manual provided criteria for analyzing human movement in terms of: 1) the amount of energy required to perform the activity; this was a seven point scale, going

from weakest to the most vigorous kinds of movement; 2) the degree of social interaction indicated by the activity; 3) the extent to which the M expresses approach or avoidance behavior; 4) the degree of conflict which occurs in perceptions.

Thetford (1952) used Zubin's scale on the Rorschach. His subjects were 170 normal children and 50 schizophrenic children, divided into three age groups: young children (6-9), prepubescent (10-13), and adolescent (14-17). In general, the schizophrenic group had a higher quantity of M, and more poor form level in their M percepts. Categorized in terms of flexor, extensor, static and ambivalent types of M, the youngest group of schizophrenics gave more vigorous and less static M than normals; this trend was reversed for the adolescent groups. For evaluated energy, as measured on the Zubin scale in terms of mean quantity of energy per response, there was a tendency in all three age groups for normals to produce a higher quantity of energy in their Ms than the schizophrenics. Despite a lack of statistical significance, Thetford's study is encouraging in that it suggests that the energy level of Rorschach movement responses may vary in a consistent and meaningful way with developmental level and pathological condition.

Richards and Lederman (1956) gave the Levy Movement Blots to 66 handicapped children, and rated the energy level of their M responses. They used a seven-point scale quite similar to Zubin's; energy was scaled from minimal (sleeping,

lying down) to maximal (jumping, climbing, doing somersaults). The energy levels of those children handicapped after infancy were significantly higher than those whose handicaps dated from infancy. The handicapped group, pupil-patients at the Illinois Children's Hospital School in Chicago, had been somewhat selected on the basis of educability and seriousness of handicap. Therefore, the study's conclusions are valid only insofar as the handicapped group can be considered as truly representative of the handicapped population.

Piotrowski and Schreiber (1952) gave repeated Rorschach examinations to non-psychotic patients during the course of therapy. Group A, composed of thirteen patients, received psychoanalytically oriented psychotherapy. Group B, ten patients, were given sporadic therapy, mainly of a supportive kind. At least three Rorschach administrations were given to all patients, at the beginning of therapy, in the middle, and at the termination. Group A showed gradual qualitative changes in M which were interpreted as improvement in the quality of the M: a tendency for the M to become freer and more expansive, and a tendency for the M figures to change in the direction of similarity to the patient, i.e., to resemble the patient more in terms of sex, age, etc. These changes were directly related to the patients' clinical improvement. Of course, these results are based on a small, highly selected sample. Nonetheless, the results suggest that the quality of the M response changes only gradually and is

related to personality change, and thus it would seem important to consider qualitative as well as quantitative features of M.

In general, the reliability coefficients for Rorschach movement scores reported in existing studies tend to be varying and contradictory. Thus it cannot be concluded with any degree of certainty that movement responses are reliable enough for individual prediction. Nor can it be concluded that they are definitely unreliable. There is some evidence tending to suggest that the quality of M is more stable than its quantity. There is further evidence that the amount of activity in a movement response reflects stable personality characteristics, and thus that reliability of movement responses would be higher if strength of movement were taken into consideration, and scores differentially weighted on that basis. The only strength of movement scale that existed before the present study, Zubin's, has several shortcomings. First, it was built only for human movement, and provides no basis for weighting animal and inanimate movement. Second, for maximum accuracy in measuring Rorschach responses, the scale should represent the range of actual Rorschach responses. Since Zubin's scale was designed for the Levy Movement Blots, it does not fulfill this requirement. In addition, Zubin's is simply a graphic rating scale, with points determined on a a priori basis. A more refined, empirically-derived scale is needed.

Such a scale was constructed for this study. Reliability coefficients were obtained for the different kinds of movement scores, taking into account the total number of responses, and the reliability of the traditional Rorschach scores was compared with the reliability of the weighted scores.

Method

Construction of the Epstein-Lambert Scale

The first step in the construction of the Epstein-Lambert scale was to collect at least 50 non-repetitive examples of human, animal and inanimate movement. A total of 71 Rorschach protocols were used for this purpose, the sample of subjects representing adults of both sexes, ranging in age from 16-65, varying in education and occupation. These 71 records yielded the requisite numbers of M and FM, and about 40 m. In order to get a total of 50 m responses with non-repetitive content, it was necessary to include about 10 examples of inanimate movement from those recorded in Beck (1950) and Klopfer (1954). Thus the sample, while not random, is a reasonably varied one, and inclusive of a wide range of Rorschach movement responses.

The second step was to Q sort the three sets of movement responses according to energy level. The use of the Q sort normalized the distribution of the items relative to each other. The responses were sorted into nine columns, with number of items in each column as follows:

3 5 7 9 13 9 7 5 3

Weights ranged from 1, minimal, doubtful expression of energy, to 9, extremely intense expression of energy. Modification of basic weights was made by considering:

1. Vividness of percept -- High emotional intensity or vividness of reaction received a higher weight than a percept of low emotional intensity.
2. Partial rejection of percept -- Attempts at rejection of the perception lowered the weight.
3. Area covered -- A response to the whole blot received higher weighting than a response to a part.
4. Abstraction -- Symbolic views or pictures received lower weights than direct representations.
5. Time -- Ongoing action received higher weighting than past or potential action.

Modification could change scores up to three points. The data were 4 sorted independently three times: once by a psychology department faculty member, and twice by clinical psychology graduate students. Interscorer agreement was then measured by correlation coefficients, which are presented below:

Scorers	Human Movement	Animal Movement	Inanimate Movement
S.K. and S.E.	.89	.80	.82
S.K. and E.L.	.88	.78	.81

After the correlations were computed, discrepancies were resolved by discussion. A set of examples of responses representative of each point on the scale was drawn up, to serve as criteria for scoring. (See Appendix for these scaling points.)

Subjects

Seventy-two double sets of protocols were collected, each double set consisting of one complete Rorschach and one Behn-Rorschach. The procedure was set to find maximum reliability coefficients by making the subject sample as heterogeneous as possible and by limiting the time between tests. The subject sample varied in occupation, including teachers, students, business men, housewives and laborers. The subjects' educational levels ranged from elementary school to post-graduate education. Personality adjustment varied from normal to psychotic. Age levels from sixteen upwards were represented.

Procedure

Each subject was given the Rorschach and the Behn-Rorschach in two sessions. The first session consisted of odd-numbered Rorschach cards and even-numbered Behn cards, the second session of odd Behns and even Rorschachs. The two sessions were separated by not less than one hour and not more than twenty-four hours. At least an hour was required between tests to reduce fatigue effects. The upper limit of 24 hours was selected to limit actual changes in personality. For most of the subjects, the 24 hour interval between sessions was used.

The testing procedure was the standard one used for Rorschach administrations. Only the movement responses to the blots were scored. A weighted score of energy level was

assigned to each response. To minimize scoring bias, a systematic order of scoring was followed. The Session I protocols collected each week were put into one envelope, the Session II protocols for that week into another. Scoring of a group of Session I records was followed by scoring of a group of Session II records for a different week. This procedure insured that each S's two records were scored independently.

The 72 subjects were divided into three groups of 24 each on the basis of total number of responses during the first testing session. Those Ss giving 0-23 responses during the first session were put in the low response group; those with 24-30 responses in the medium group; those with 31 and over in the high group. Too many Ss, however, had scores of 24 and 31 to permit even divisions; in these cases Ss were assigned to response groups on the basis of combined number of responses for both sessions. The range for the first session for the low group was 11-24, for the medium group 24-31, for the high group 31-82. The respective means were 19.33, 26.80, and 40.25.

Table 1 shows the experimental design. Each S received the Rorschach odd-numbered cards and Behn even-numbered during the first session, and Rorschach even-numbered cards and Behn odd numbered cards during the second session. Thus there is a confounding of order and odd-even cards. A significant difference between sessions could indicate order differences, or could mean that combined Rorschach odd and

Table 1

Experimental design

Total R	Session I		Session II	
	Rorschach cards	Behn cards	Rorschach cards	Behn cards
11-24	I	II	II	I
	III	IV	IV	III
	V	VI	VI	V
	VII	VIII	VIII	VII
	IX	X	X	IX
24-31	I	II	II	I
	III	IV	IV	III
	V	VI	VI	V
	VII	VIII	VIII	VII
	IX	X	X	IX
31-82	I	II	II	I
	III	IV	IV	III
	V	VI	VI	V
	VII	VIII	VIII	VII
	IX	X	X	IX

Behn even cards are different from combined Rorschach even and Behn odd cards. Thus session differences must be interpreted with caution, as must all interactions into which the sessions effect enters. Since this confounding of odd-even cards and order makes interpretation difficult, the .01 level of significance is set for interpretation of within subjects effects.

Results

Separate analyses of variance were performed for the Rorschach human movement, animal movement, and inanimate movement scores, and for the same three types of energy level scores. Additional analyses were done for a combined Rorschach M + FM + m, and for a total energy scores, making a total of eight analyses. The variables were: response-groups, sessions, and tests.

Two reliability coefficients were derived from each analysis of variance, the first reflecting consistency in performance from first to second administrations across tests, the second reflecting consistency in performance from Behn to Rorschach tests across sessions. The coefficients were computed from mean squares, by applying formulae given by Haggard (1958). The formula for reliability for sessions across tests is:

$$[1] \quad r_{\text{sessions}} = \frac{\sigma^2_{\underline{S}s/R}}{\sigma^2_{\underline{S}s/R} + \sigma^2_{\underline{S}s} \times Ses/R + \sigma^2_e}$$

where $\sigma^2_{\underline{S}s/R}$ = total variance due to individual differences,

$\sigma^2_{\underline{S}s} \times Ses/R$ = variance due to individual differences from session to session,

σ^2_e = error variance, i.e., variance not accounted for by experimental variables.

The formula for reliability for tests across sessions is:

$$[2] \quad r_{test} = \frac{\sigma^2_{\underline{S}s/R}}{\sigma^2_{\underline{S}s/R} + \sigma^2_{\underline{S}s} \times T/R + \sigma^2_e}$$

where $\sigma^2_{\underline{S}s} \times T/R$ = variance due to individual differences from test to test.

Human Movement

Table 2 presents the analysis of variance for human movement responses and the reliability coefficients derived from it. Table 3 gives mean numbers of human movement responses.

The F ratio for response groups is 5.73, which is significant at the .01 level. The means, in increasing order of total R, are 2.40, 2.59, and 4.27, indicating a direct relationship between total number of responses and number of M responses.

The F ratio for sessions is 40.36, which is significant at the .01 level. Table 3 shows a sharp drop in mean number of M from first to second session, from 4.21 to 2.14.

The F ratio for tests is 7.05, which is significant at the .01 level. Table 3 shows more M responses on the Rorschach than on the Behn, with means of 3.53 and 2.82 respectively.

The sessions x tests interaction has an F ratio of 9.46, which is significant at the .01 level. The mean number of M responses to the Rorschach odd cards in the first session was 2.47, and to the Rorschach even cards in the second session was 1.06. The mean for the Behn even cards in the first

Table 2

Analysis of variance of Rorschach human movement scores (M)

Source	df	SS	MS	F
Total	287	729.83		
Between Subjects	71	322.08		
Response Groups	2	45.86	22.93	5.73**
Ss/R	69	276.22	4.00	3.60***
Within Subjects	216	407.75		
Sessions	1	77.09	77.09	40.36***
Tests	1	9.03	9.03	7.05**
Ses x T	1	10.50	10.50	9.46**
R x Ses	2	8.69	4.35	2.28
R x T	2	1.75	0.88	1.37
R x Ses x T	2	4.20	2.10	1.81
Ss x Ses/R	69	131.47	1.91	1.72**
Ss x T/R	69	88.47	1.28	1.15
Ss x Ses x T/R	69	76.55	1.11	

** Significant at .01 level

*** Significant at .001 level

$r_{sessions} = 0.32$

$r_{test} = 0.38$

Table 3

Mean numbers of M responses

Total R	Session I			Session II			Grand Total
	Rorschach odd cards	Behn even	Total	Rorschach even cards	Behn odd	Total	
11-24	2.04	1.21	3.25	0.75	0.79	1.54	2.40
24-31	2.08	1.50	3.58	1.29	0.83	2.12	2.59
31-82	3.29	2.50	5.79	1.13	1.63	2.76	4.27
All Ss	2.47	1.74	4.21	1.06	1.08	2.14	3.18

session was 1.74, for the Behn odd cards in the second session 1.08. These means show that there was a much greater drop in number of M responses in the second session for the Rorschach than for the Behn. This difference could be due to either the differences between the individual blots in the odd-even division of the Rorschach and Behn or to an order effect.

The F ratio for Bs/R is 4.00, which is significant at the .001 level. This indicates that M scores do reflect individual differences to a highly significant degree, but does not indicate the degree of reliability. This can be determined by turning to the reliability coefficients derived from the mean squares. A reliability coefficient of .32 is found for reliability between sessions and a coefficient of .38 for reliability between tests.

Animal Movement

Table 4 presents the analysis of variance for animal movement and the reliability coefficients derived from it. Table 5 gives mean numbers of FM responses.

The F ratio for response groups is 4.87, which is significant at the .05 level. The means, in increasing order of total R, are 3.00, 4.27 and 4.63, indicating a direct relationship between total number of responses and number of FM responses.

The F ratio for sessions is 13.54, which is significant at the .001 level. The means are 3.42 for the first session

Table 4

Analysis of variance of Rorschach animal movement scores (FM)

Source	df	SS	MS	F
Total	287	670.91		
Between Subjects	71	283.16		
Response Groups	2	35.05	17.53	4.87*
Ss/R	69	248.11	3.60	1.96***
Within Subjects	216	387.75		
Sessions	1	21.67	21.67	13.54***
Tests	1	9.75	9.75	8.63**
Ses x T	1	21.70	21.70	11.70**
R x Ses	2	6.30	3.15	1.97
R x T	2	0.49	0.25	0.22
R x Ses x T	2	12.54	6.27	3.40*
Ss x Ses/R	69	110.55	1.60	0.87
Ss x T/R	69	78.04	1.13	0.61
Ss x Ses x T/R	69	127.01	1.84	

* Significant at .05 level

** Significant at .01 level

*** Significant at .001 level

$r_{sessions} = 0.20$

$r_{test} = 0.28$

Table 5

Mean numbers of FM responses

Total R	Session I			Session II			Grand Total
	Rorschach odd cards	Behn even	Total	Rorschach even cards	Behn odd	Total	
11-24	0.88	1.17	2.05	1.79	2.17	3.96	3.00
24-31	1.29	2.63	3.92	2.50	2.13	4.63	4.27
31-82	1.58	2.71	4.29	2.75	2.21	4.96	4.63
All Ss	1.25	2.17	3.42	2.35	2.17	4.52	3.97

and 4.51 for the second session. This difference is in the opposite direction from that of M responses, which were more numerous during the first session than during the second session.

The F ratio for tests is significant at the .01 level. The means are 3.60 for Rorschach cards and 4.33 for Behn cards, indicating more FM responses are elicited by the Behn than by the Rorschach. This difference is in the opposite direction from that of M responses, which were more numerous for Rorschach cards than for Behn cards.

The sessions by tests interaction is significant at the .01 level. The mean number of FM responses to Rorschach odd cards in the first session is 1.25, and to the Rorschach even cards in the second session is 2.35. The mean for Behn even cards in the first session is 2.17, for the Behn odd cards in the second session 2.17. These means show that there was a rise in the number of FM responses given to the Rorschach even cards in the second session, while the number of FM responses given to Behn cards remained constant over the two sessions. This could be due to either the difference between the Rorschach odd and even cards, or to an order effect.

The F ratio for Ss/T is significant at the .001 level. This indicates that FM scores reflect individual differences to a highly significant degree. The reliability coefficients are .20 for sessions and .28 for tests.

Inanimate Movement

Table 6 presents the analysis of variance for inanimate movement and the reliability coefficients derived from it. Table 7 gives mean numbers of these m responses.

Table 6 indicates that the F ratio for response groups is 4.65, which is significant at the .05 level. The means, in increasing order of total R, are 1.31, 1.43 and 2.40, indicating a direct relationship between total number of responses and number of m responses.

The F ratio for sessions is significant at the .01 level. The means are 2.23 for Session I and 1.09 for Session II, showing a greater number of m responses in the first session than in the second.

The F ratio for tests is significant at the .01 level. The mean number of m for the Rorschach is 1.39, for the Behn 2.14, indicating that the Behn elicited more m responses than the Rorschach.

The F ratio for subjects is significant at the .001 level. This indicates that m scores measure individual differences to a significant degree. A reliability coefficient of .22 was found for sessions, and a coefficient of .23 for tests.

Combined Movement Scores

Table 8 presents the analysis of variance for combined movement scores and the reliability coefficients derived from it. Table 9 gives mean combined energy scores.

Table 6

Analysis of variance of Rorschach

inanimate movement scores (m)

Source	df	SS	MS	F
Total	287	317.99		
Between Subjects	71	128.49		
Response Groups	2	15.26	7.63	4.65*
Ss/R	69	113.23	1.64	2.16***
Within Subjects	216	189.50		
Sessions	1	15.13	15.13	19.15**
Tests	1	10.13	10.13	14.07**
Ses x T	1	2.34	2.34	3.08
R x Ses	2	3.52	1.76	2.23*
R x T	2	0.39	0.20	0.28
R x Ses x T	2	0.55	0.28	0.37
Ss x Ses/R	69	54.84	0.79	1.04
Ss x T/R	69	49.98	0.72	0.95
Ss x Ses x T/R	69	52.61	0.76	

* Significant at .05 level

** Significant at .01 level

*** Significant at .001 level

$r_{sessions} = 0.22$

$r_{test} = 0.23$

Table 7

Mean numbers of m responses

Total N	Session I			Session II			Grand Total
	Rorschach odd cards	Behn even	Total	Rorschach even cards	Behn odd	Total	
11-24	0.54	1.08	1.63	0.50	0.50	1.00	1.31
24-31	0.63	1.25	1.88	0.54	0.75	1.29	1.43
31-82	1.33	1.83	3.17	0.63	1.00	1.63	2.40
All Ss	0.83	1.39	2.22	0.56	0.75	1.31	1.71

Table 8

Analysis of variance of combined Rorschach movement scores (M + FM + m)

Source	df	SS	MS	F
Total	287	2069.11		
Between Subjects	71	1220.11		
Response Groups	2	269.84	134.92	9.80***
Ss/R	69	950.27	13.77	4.72***
Within Subjects	216	849.00		
Sessions	1	64.22	64.22	14.37***
Tests	1	10.12	10.12	3.50
Ses x T	1	6.13	6.13	2.10
R x Ses	2	34.09	17.05	3.81*
R x T	2	1.19	0.60	0.61
R x Ses x T	2	23.52	11.76	4.03*
Ss x Ses/R	69	308.69	4.47	1.53*
Ss x T/R	69	199.69	2.89	0.99
Ss x Ses x T/R	69	201.35	2.92	

* Significant at .05 level

*** Significant at .001 level

$r_{sessions} = 0.42$

$r_{test} = 0.48$

Table 9

Mean numbers of H + (H) + h responses

Total R	Session I			Session II			Grand Total
	Rorschach odd cards	Behn even	Total	Rorschach even cards	Behn odd	Total	
11-24	3.54	3.38	6.92	2.88	3.46	6.33	6.63
24-31	3.96	5.38	9.33	4.33	3.71	8.04	8.69
31-82	6.25	7.00	13.25	4.58	4.88	9.46	11.35
All Ss	4.58	5.25	9.83	3.93	4.01	7.94	8.89

The F ratio for response groups is significant at the .001 level. The means, in increasing order of total R, are 6.63, 8.69 and 11.35, indicating a direct relationship between total number of responses and total movement responses.

The F ratio for sessions is significant at the .001 level. Means of 9.83 in the first session and 7.94 in the second session indicate that there were less movement responses in the second session than in the first.

The F ratio for subjects is significant at the .001 level, which means that a total movement score measures individual differences to a significant degree. Reliability coefficients of .42 for sessions and .48 for tests were found. These are higher than any coefficients obtained for M, FM or m alone.

Human Energy

Table 10 presents the analysis of variance for human energy scores and the reliability coefficients derived from it. Table 11 gives mean human energy scores.

The F ratio for response groups is significant at the .05 level. The means, in increasing order of total R, are 10.82, 12.86 and 18.17, indicating a direct relationship between total number of responses and human energy scores.

The F ratio for sessions is significant at the .001 level. The mean energy score for the first session is 19.14, for the second session 8.75, showing a sharp drop in human energy scores for the second session.

Table 10

Analysis of variance of human energy level scores (EH)

Source	df	SS	MS	F
Total	287	15081.78		
Between Subjects	71	5958.28		
Response Groups	2	691.80	345.90	4.53*
Ss/R	69	5266.48	76.33	3.08***
Within Subjects	216	9123.50		
Sessions	1	1942.72	1942.72	52.25***
Tests	1	130.68	130.68	3.67
Ses x T	1	82.35	82.35	3.32
R x Ses	2	167.22	83.61	3.38
R x T	2	51.09	25.55	0.72
R x Ses x T	2	15.00	7.50	0.30
Ss x Ses/R	69	2564.56	37.17	1.50
Ss x T/R	69	2457.73	35.62	1.44
Ss x Ses x T/R	69	1712.15	24.81	

* Significant at .05 level

*** Significant at .001 level

$r_{sessions} = 0.29$

$r_{test} = 0.30$

Table 11

Mean human energy scores (EH)

Total R	Session I			Session II			Grand Total
	Rorschach odd cards	Behn even	Total	Rorschach even cards	Behn odd	Total	
11-24	8.88	6.29	15.17	3.25	3.21	6.46	10.82
24-31	9.79	6.96	16.75	5.46	3.50	8.96	12.86
31-82	13.67	11.83	25.50	4.83	6.00	10.83	18.17
All Ss	10.78	8.36	19.14	4.51	4.24	8.75	13.95

The F ratio for subjects is significant at the .001 level, which indicates that human energy scores measure individual differences to a significant extent. Reliability coefficients of .29 for sessions and .30 for tests were found. These are slightly lower than the coefficients of .32 for sessions and .38 for tests found for human movement scores. Thus it seems that weighting human movement scores for energy level does not improve their reliability.

Animal Energy

Table 12 presents the analysis of variance for animal energy and the reliability coefficients derived from it. Table 13 gives mean animal energy scores.

The F ratio for response groups is significant at the .01 level. The means in increasing order of total R, are 11.88, 19.17 and 19.50, indicating a direct relationship between total number of responses and animal energy scores.

The F ratio for tests is significant at the .01 level. The mean EA score for the Rorschach is 14.31, for the Behn 18.89, showing higher EA scores for Behn cards.

The sessions by tests interaction is significant at the .001 level. The mean EA score for Rorschach odd cards in the first session was 5.32, and for Rorschach even cards in the second session 9.49. The mean for Behn even cards in the first session was 9.74, and for Behn odd cards in the second session 9.15. These means show that there was a rise in EA scores for Rorschach even cards in the second session, while

Table 12

Analysis of variance of animal energy scores (EA)

Source	df	SS	MS	F
Total	287	13802.32		
Between Subjects	71	5773.82		
Response Groups	2	891.36	445.68	6.86**
Ss/R	69	4882.46	64.96	1.96***
Within Subjects	216	8028.50		
Sessions	1	231.13	231.13	6.70*
Tests	1	300.13	300.13	11.88**
Ses x T	1	406.12	406.12	12.27***
R x Ses	2	88.58	44.29	1.21
R x T	2	8.33	4.17	0.16
R x Ses x T	2	449.34	224.67	6.79
Ss x Ses/R	69	2517.79	36.49	1.10
Ss x T/R	69	1743.04	25.26	0.76
Ss x Ses x T/R	69	2284.04	33.10	

* Significant at .05 level

** Significant at .01 level

*** Significant at .001 level

$r_{sessions} = 0.19$

$r_{test} = 0.21$

Table 13

Mean animal energy scores (EA)

Total R	Session I			Session II			Grand Total
	Rorschach odd cards	Behn even	Total	Rorschach even cards	Behn odd	Total	
11-24	4.08	4.58	8.67	6.17	8.92	15.08	11.88
24-31	5.88	11.63	17.50	11.25	9.58	20.83	19.17
31-82	6.00	13.00	19.00	11.04	8.96	20.00	19.50
All Ss	5.32	9.74	15.06	9.49	9.15	18.64	16.85

the Behn odd cards showed a slight drop in the second session. These could be due to odd-even differences, or to an order effect.

The F ratio for subjects is significant at the .001 level, showing that the animal energy score measures individual differences to a significant degree. The reliability coefficients found were .19 for sessions and .21 for tests. Comparing them with the coefficients of .20 for sessions and .28 for tests found for FM, it may be seen that EA scores are slightly less reliable than FM scores.

Inanimate Energy

Table 14 presents the analysis of variance for inanimate energy scores and the reliability coefficients derived from it. Table 15 gives mean animal energy scores.

The F ratio for response groups is significant at the .05 level. The means, in increasing order of total R, are 6.57, 8.19, and 11.98, indicating a direct relationship between total number of responses and inanimate energy scores.

The F ratio for sessions is significant at the .01 level. Table 14 shows a drop in mean EO score from first to second session, from 10.85 to 6.99.

The F ratio for tests is significant at the .01 level. The mean EO score for the Rorschach cards was 7.13, for the Behn cards 10.68. This indicates higher EO scores given to the Behn than to the Rorschach cards.

The F ratio for subjects is significant at the .001

Table 14

Analysis of variance of inanimate energy scores (EO)

Source	df	SS	MS	F
Total	287	9069.32		
Between Subjects	71	4306.32		
Response Groups	2	373.22	186.61	3.79*
Ss/R	69	2933.10	49.18	2.80***
Within Subjects	216	4763.00		
Sessions	1	264.50	264.50	11.58**
Tests	1	227.56	227.56	11.40**
Ses x T	1	13.34	13.34	0.76
R x Ses	2	46.64	23.32	1.02
R x T	2	30.67	15.34	0.77
R x Ses x T	2	15.22	7.61	0.43
Ss x Ses/R	69	1575.86	22.84	1.30
Ss x T/R	69	1377.77	19.97	1.14
Ss x Ses x T/R	69	1211.44	17.56	

* Significant at .05 level

** Significant at .01 level

*** Significant at .001 level

$r_{sessions} = 0.28$

$r_{tests} = 0.30$

Table 15

Mean inanimate energy scores (EO)

Total R	Session I			Session II			Grand Total
	Rorschach odd cards	Behn even	Total	Rorschach even cards	Behn odd	Total	
11-24	3.17	4.50	7.67	2.50	2.92	5.42	6.57
24-31	3.38	6.42	9.79	2.75	3.83	6.58	8.19
31-82	6.38	8.63	15.01	3.21	5.75	8.96	11.98
All Ss	4.31	6.51	10.82	2.82	4.23	7.05	8.91

level, showing that the FC score measures individual differences to a significant degree. The reliability coefficients found were .28 for sessions and .30 for tests. Comparing them with the reliability coefficients of .22 for sessions and .23 for tests found for m, it may be seen that weighted inanimate energy scores are slightly more reliable than number of inanimate movement scores.

Total Energy

Table 16 presents the analysis of variance for total energy scores and the reliability coefficients derived from it. Table 17 gives mean total energy scores.

The F ratio for response groups is significant at the .001 level. The means, in increasing order of total R are 29.13, 40.09 and 49.65, indicating a direct relationship between total number of responses and total energy score.

The F ratio for sessions is significant at the .001 level. The mean for the first session is 44.93, for the second session 34.31, showing a drop in total energy score from the first to the second session.

The F ratio for subjects is significant at the .001 level, showing that combined energy scores measure individual differences to a significant degree. The reliability coefficients found were .41 for sessions and .48 for tests. These coefficients are higher than any obtained for FM, FA or FC alone. They are almost identical with the coefficients of .42 for sessions and .48 for tests found for combined movement

Table 16

Analysis of variance of combined energy scores (EH + EA + EO)

Source	df	SS	MS	F
Total	287	47186.50		
Between Subjects	71	25772.75		
Response Groups	2	5061.05	2530.53	8.43***
Ss/R	69	20711.70	300.17	4.41***
Within Subjects	216	21413.75		
Sessions	1	2032.03	2032.03	20.73***
Tests	1	432.67	432.67	5.02*
Ses x T	1	163.51	163.51	2.40
R x Ses	2	763.90	381.95	3.90*
R x T	2	125.72	62.86	0.73
R x Ses x T	2	494.25	247.13	3.63*
Ss x Ses/R	69	6764.32	98.03	1.44*
Ss x T/R	69	5945.86	86.17	1.27
Ss x Ses x T/R	69	4691.49	67.99	

* Significant at .05 level

*** Significant at .001 level

$r_{sessions} = 0.41$

$r_{tests} = 0.48$

Table 17

Mean combined energy scores (ZK + CA + BO)

Total N	Session I			Session II			Grand Total
	Rorschach odd cards	Behn even	Total	Rorschach even cards	Behn odd	Total	
11-24	16.13	15.38	31.51	11.63	15.13	26.76	29.13
24-31	19.08	24.71	43.79	19.33	17.04	36.37	40.09
31-82	26.25	33.25	59.50	19.08	20.71	39.79	49.65
All Ss	20.49	24.45	44.94	16.68	17.63	34.31	39.62

scores. Thus it seems that weighting combined movement scores for energy level has no appreciable effect on their reliability.

Summary of Results

Table 18 presents a summary of the analyses of variance.

For all analyses, response groups differed significantly: the higher the number of total responses, the higher movement or energy score that was found.

The sessions variable was significant at or above the .01 level for all analyses except EA. All scores except FM and EA were lower during the second session. For FM higher scores were found during the second session than during the first.

The tests variable was significant at the .01 level for all analyses except EM and the two combined scores. For M, more responses were given to Rorschach than Behn cards. For FM and EA, and for m and CO, higher scores were found for the Behn cards than for the Rorschach.

The sessions by tests interaction was significant at or above the .01 level for M, FM and EA. Many more M responses were given to the Rorschach odd cards than to the Behn even cards during the first session, whereas there was a slight difference in favor of the Behn odd cards during the second session. For both FM and EA a difference in favor of the Behn even cards during the first session, and a smaller difference in favor of the Rorschach even cards during the

Table 18

Summary of analysis of varience

Source	M	FM	m	M + FM + m	EH	EA	EO	EH + EA + EO
Between Ss								
R groups	**	*	*	***	*	**	*	***
Ss/R	***	***	***	***	***	***	***	***
Within Ss								
Sessions	***	***	**	***	***	*	**	***
Tests	**	**	**			**	**	*
Ses x T	**	**				***		
R x Ses			*	*				*
R x T								
R x Ses x T		*		*				*
Ss x Ses/R	**			*				*
Ss x T/R								
Ss x Ses x T/R								

* Significant at the .05 level

** Significant at the .01 level

*** Significant at the .001 level

second session were found.

Table 19 presents the reliability coefficients derived from the analyses of variance. The magnitude of the reliability coefficients can be seen to vary from .19 to .48. Table 19 indicates that weighting for energy level in human and animal percepts slightly reduces reliability, but that weighting for energy level in inanimate object percepts slightly raises reliability. There is no test of whether these differences are significant. For the combined score, energy level weighting makes no difference.

The coefficients obtained are uniformly low. Although the combined scores are more reliable than their components, certainly none are sufficiently reliable to permit individual prediction.

Table 19

Reliability coefficients derived

from analyses of variance

	M	FM	fm	M + FM + fm
$r_{sessions}$	.32	.20	.22	.42
r_{test}	.38	.28	.23	.48

	EH	EA	EO	EH + EA + EO
$r_{sessions}$	.29	.19	.28	.41
r_{test}	.30	.21	.30	.48

Discussion

The purpose of this study was to test the reliability of Rorschach movement responses and movement responses weighted for energy level under conditions designed to maximize reliability. The reliability coefficients found varied from .19 to .43. These coefficients were found to be significant measures of individual differences, but the degree of individual prediction possible from session to session or from test to test is very limited.

The coefficients for sessions and test respectively of .32 and .38 for M, .20 and .28 for FM, and .22 and .23 for m are lower than those found in most previous reliability studies. Vernon (Hertz, 1934) found a reliability of .62 for M; Hertz (1934) found a reliability coefficient of .76 for M. Thornton and Guilford (1936) found reliability coefficients of .919 and .768 for M. These studies were methodologically faulty, however. They used the split-half technique, which is not a suitable one for testing Rorschach reliability (Cronbach, 1954). As discussed in the introduction, some of their statistical techniques were unsound, and they did not control for total number of responses.

Eichler (1951) found coefficients of .82 for M and .46 for FM under test-retest conditions, and coefficients of .50 and .51 with parallel-form testing. However, Eichler did not control for total number of responses. Singer (1952), using

the Behn as a parallel form, found coefficients of .39 for M% and .81 for number of M. However, Singer's experimental group consisted of only 10 Rs, to which were added 23 cases from Zulliger's manual. The scoring for the two groups could not be completely equated. Also, he did not control for total number of responses. Schwartz and Yates (1957) found Rorschach-Behn parallel form reliability coefficients of .45 for M, .58 for FM and .72 for m. The study is limited by being based on only 12 subjects. Also, they controlled for total R by requiring a fixed number of responses per card, which makes the task somewhat different from the standard Rorschach task, and thus limits the possible generalization to the usual Rorschach.

Epstein, Nelson and Tanofsky (1957) constructed 100 inkblots which were assigned randomly to 10 sets of 10 cards each. They found reliability coefficients of .42 for M, .27 for FM, and .23 for m. These agree fairly well with the coefficients found in the present study. They also found that a combined movement score (M + FM + m) was more reliable than any of its components, with a coefficient of .53. The task differed from the standard Rorschach, in the kind of inkblots used and in the requirement of a fixed number of responses per card. However, their results are consistent with the results of the present study, which found a combined movement score to have a reliability coefficient of .42 for sessions and .48 for tests.

In general, then, previous studies have found higher reliabilities than the present study. Many of these previous studies, however, used faulty methodology and lack necessary controls, which can be assumed to have inflated their reliability coefficients. From the present study it may be concluded that the movement score does significantly measure some individual difference, but that this measurement is not reliable enough to permit individual prediction.

It was hypothesized that weighting scores for energy level would make them more reliable. This hypothesis was not substantiated. A combined energy score had higher reliability than any of its components, which indicates that the scores do measure something in common, which might correspond to an overall energy level. However, the measurement is not reliable enough to permit individual prediction.

Of incidental interest are the patterns of session and test differences. Human and inanimate movement and the corresponding energy scores dropped during the second session. Animal movement rose during the second session. It is difficult to interpret these session differences, because of the confounding of odd-even differences and order effects. It may be that the session differences simply indicate that the combination of Rorschach odd and Behn even cards elicits different responses than the combination of Rorschach even and Behn odd cards. A second possibility is that order effects are responsible for the pattern. One might speculate

that fatigue effected the second session. Human movement is said to reflect better control and integration of fantasy than animal movement (Klopfer, 1954), which fatigue can be presumed to impair. The existence of fatigue is supported by the drop in combined movement and energy scores during the second session. A third possibility is that both odd-even differences and order effects contributed to the session differences.

The Rorschach cards elicited more human movement than the Behn cards, the Behn more animal and inanimate movement and energy than the Rorschach. Parts of this pattern have been found by some previous investigators. Eichler (1951) and Tulliger (1956) reported that the Behn elicited more animal content and animal movement than the Rorschach. Eichler (1951) also found a tendency for the Behn to elicit less human content and human movement than the Rorschach. These differences should be taken into consideration in future studies with the Rorschach and Behn parallel forms, and in testing patients with the two forms.

Summary

The purpose of this study was to test the reliability of Rorschach movement responses, obtaining maximal estimates of reliability. This was done by using analysis of variance to eliminate incidental sources of variance such as total R and session effects, by keeping time between tests relatively short, and by using a heterogeneous subject sample. A secondary purpose of the study was to ascertain if the reliability of Rorschach movement responses could be improved by weighting the scores for energy level. A new scale for weighting energy level of Rorschach responses was constructed. Each of 72 Ss was given the Rorschach and Behn-Rorschach in two sessions. The first session consisted of the five odd-numbered Rorschach cards and the five even-numbered Behn cards, the second of odd-numbered Behn cards and even-numbered Rorschach cards. The Ss were divided into three groups on the basis of total number of responses.

Analyses of variance were carried out on Rorschach M, FM and m scores, the corresponding energy level scores, and for a combined Rorschach score and a total energy score. Reliability coefficients for sessions and for tests were derived from these analyses.

The major findings may be summarized as follows:

1. The reliability coefficients found were uniformly low, ranging from .19 to .48, none high enough to

serve as a basis for individual prediction.

2. Weighting for energy level did not improve reliability.
3. The combined movement score and combined energy score were more reliable than any of their components.
4. For all analyses, there was a direct relationship between total R and the score measured. This indicates the importance of controlling for total R in reliability studies.
5. All scores dropped significantly during the second session except animal movement and energy. Animal movement and energy rose during the second session.
6. The Behn elicited more animal and inanimate movement and energy than the Rorschach. The Rorschach elicited more human movement than the Behn.

References

Beck, S.J. Rorschach's test. Vol. 1. Basic processes. New York: Grune and Stratton, 1950.

Cronbach, L. Statistical methods applied to Rorschach scores: a review. Psychol. Bull., 1949, 46, 393-429.

Eichler, R.M. A comparison of the Rorschach and Behn ink-blot tests. J. consult. Psychol., 1951, 15, 185-189.

Epstein, S., Nelson, J., & Tanofsky, R. Responses to ink-blots as measures of individual differences. J. consult. Psychol., 1957, 21, 211-215.

Fosberg, I. An experimental study of the reliability of the Rorschach psychodiagnostic technique. Rorschach Res. Exch., 1941, 5, 72-84.

Haggard, E. Intraclass correlation and the analysis of variance. New York: The Dryden Press, Inc., 1958.

Hertz, M. Reliability of the Rorschach inkblot test. J. appl. Psychol., 1934, 18, 461-477.

Holzberg, J., & Wexler, M. The predictability of schizophrenic performance on the Rorschach test. J. consult. Psychol., 1950, 14, 395-399.

Kelley, D., Margulies, H., & Barrera, S. The stability of the Rorschach method as demonstrated in electric convulsive therapy cases. Rorschach Res. Exch., 1941, 5, 35-43.

Klopfer, B., Ainsworth, M., Klopfer, W., & Holt, R. Developments in the Rorschach technique. Yonkers, N.Y.: World Book, 1954.

Meltzoff, J., Singer, S., & Korchin, S. Motor inhibition and Rorschach human movement responses. J. Pers., 1953, 21, 400-410.

Piotrowski, Z., & Schreiber, M. Rorschach perceptanalytic measurement of personality changes during and after intensive psychoanalytically oriented psychotherapy. In: G. Bychowski & J.L. Despert (Eds.), Specialized techniques in psychotherapy. New York: Basic Books, 1952. Pp. 337-361.

Richards, T., & Lederman, E. A study of action in the fantasy of physically handicapped children. J. clin. Psychol., 1956, 12, 188-190.

Rorschach, H. Psychodiagnostics. (3rd ed.) New York: Grune and Stratton, 1951.

Schwartz, F., & Yates, S. Behn-Rorschach and Rorschach under standard and stress conditions. J. consult. Psychol., 1957, 21, 335-338.

Singer, J. The Behn-Rorschach inkblots: a preliminary comparison with the original Rorschach series. J. proj. Tech., 1952, 16, 238-245.

Singer, J., Meltzoff, J., & Goldman, G. Rorschach M responses following motor inhibition and hyperactivity. J. consult. Psychol., 1952, 16, 359-364.

Singer, J., & Spohn, H. Some behavioral correlates of Rorschach's experience type. J. consult. Psychol., 1954, 18, 1-9.

Thetford, W. Fantasy perceptions in the personality development of normal and deviant children. Amer. J. Orthopsychiat., 1952, 22, 542-550.

Thornton, G., & Guilford, J. The reliability and meaning of Erlebnistypus scores in the Rorschach test. J. abnorm. soc. Psychol., 1936, 31, 324-330.

Zubin, J. Manual of projective and cognate techniques. Madison, Wis.: The College Typing Co., 1948.

Zubin, J. Failures of the Rorschach technique. J. proj. Tech., 1954, 18, 303-315.

Zulliger, . The Behn-Rorschach Test. Bern: Hans Huber, 1956.

Examples of scores for 9 items to be used in scaling Weighting

1. a grinning face
   a person asleep

2. sitting
   standing
   leaning
   peering
   an old woman holding her head in her hands

3. arms held up
   reading
   holding something
   huddling
   making faces
   hands sticking out

4. choir director without a head, directing
   talking
   bowing
   waving
   cooking
   making faces
   kissing
   sitting and sewing

5. children playing peas porridge hot
   picking up something
   climbing down a tree
   creeping
   shaking hands
   performing rites
   talking a mile a minute
   looking startled
   leaning and opposing one another

6. women hugging
   somebody diving
   people climbing a mountain
   a dancer carefully balanced
   witches flying
   trying to hold onto something
   women pulling a pot

7. running
   pushing against an obstacle
   dancing
   riding a motor cycle

8. intoxicated students whooping it up
   kicking arms and legs
   vigorous dancing, jitterbugging; dueling; calisthenics

9. enraged giant jumping up and down

Examples of scores for 99 items to be used in scaling Weighting

1. hibernating
   just sitting

2. upper part of a frog croaking
   elephants touching trunks
   opossum hanging by the tail
   hound dog peering
   baby bird with its mouth open
   cat looking at you
   collie standing -- proud, like in dog show

3. braying
   hovering
   nuzzling
   spreading of wings
   sucking a bottle

4. parasite chewing a rabbit
   animal stepping from rock to rock
   crawling up
   creeping
   holding something
   stretching

5. sea horses spraying each other
   trick seal balancing something on its nose
   digging
   climbing
   flying (bird, butterfly)
   moose swimming
   sitting up begging

6. bears playing pattycake
   beavers walking on front feet
   birds racing
   caterpillars making a toast -- with dancing bird on its nose
   dancing bears
   monkeys throwing sausages
   mice squeezing through an opening
   crabs grabbing small animal
   animal with foot stuck, trying to pull it out
   sheep jumping around
   muskrat jumping from one rock to another
   animal climbing laboriously

7. animal rearing
   animal leaning backward, pulling

8. animal flying over something
   fighting animals
   taking a leap
   running, hurrying

9. deer running for its life
   maddened animal charging

Examples of scores for m items to be used in scaling Weighting

1. animal skin being pulled tight

2. candle flame coming out of a pink holder
   smoke from a train
   ripples dying away
   rabbit with green clouds coming out of his eyes--
   symbolize mystic experience

3. blood dripping
   boat drifting
   blood spurting
   two forces pushing together
   scattering clouds
   a pleasure boat going down stream
   earth separating gradually

4. hypodermic needle drawing fluid
   flags blowing
   wind blowing ears of two bears
   marionettes bowing being pulled by strings
   ornamental gates swinging shut
   a moving sailboat

5. feeling of disintegration
   fountain flowing up
   a spinning top
   fire in a fireplace
   cherubs falling through the sky
   torches burning
   balanced rocks
   water flowing
   rays of magic leaving a magician's arm
   tides hitting a spot of land -- over time

6. bomb falling
   airplane in flight
   metal chips, flying off the anvil
   red shoes that keep on dancing
   red symbolic of fire in hell
   waterfall

7. forest fire beginning to creep up
   Old Faithful coming up and bubbling
   projectile -- has gone through center and left path in its wake
   fireworks bursting in the air

8. atomic explosion
   a bomb blasting off
   a rocket blasting off
   aerial view of wreck -- a train, burst of flame where oil tank hit

9. world spinning around so fast that everything being spilled off into the atmosphere
   a volcano erupting in all directions

Acknowledgments

The author would like to express her deepest appreciation to Dr. Seymour Epstein, who suggested the topic of the present thesis and who served as a constant source of guidance throughout its completion.

The author would also like to express her appreciation to Dr. Jerome Myers for serving on her committee and for his help with the statistical aspects of this thesis.

In addition, the author would like to thank Dr. Vincent Rogers for serving on her committee.

Approved by:

Seymour Epstein

Curt R. Rogers

D. L. Myers

Date: 12/18/61