

1979

Test-Retest Reliability of the Hand Test with the Institutionalized Elderly

Thomas J. Lundquist

Eastern Illinois University

This research is a product of the graduate program in [Psychology](#) at Eastern Illinois University. [Find out more](#) about the program.

Recommended Citation

Lundquist, Thomas J., "Test-Retest Reliability of the Hand Test with the Institutionalized Elderly" (1979). *Masters Theses*. 3167.
<https://thekeep.eiu.edu/theses/3167>

This is brought to you for free and open access by the Student Theses & Publications at The Keep. It has been accepted for inclusion in Masters Theses by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.

THESIS REPRODUCTION CERTIFICATE

TO: Graduate Degree Candidates who have written formal theses.

SUBJECT: Permission to reproduce theses.

The University Library is receiving a number of requests from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow theses to be copied.

Please sign one of the following statements:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

Date	Author
------	--------

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because _____

8/7/79

Date	Author
------	--------

TEST-RETEST RELIABILITY OF THE HAND TEST

WITH THE INSTITUTIONALIZED ELDERLY

(TITLE)

BY

Thomas J. Lundquist

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF ARTS

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY

CHARLESTON, ILLINOIS

1979

YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

Aug. 6, 1979
DATE

ADVISER

Aug 6, 1979
DATE

DEPARTMENT HEAD

TEST-RETEST RELIABILITY OF THE HAND TEST WITH
THE INSTITUTIONALIZED ELDERLY

BY

THOMAS J. LUNDQUIST

B. S. in Psych., Northland college, 1977

ABSTRACT OF A THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Psychology at the Graduate
School of Eastern Illinois University

CHARLESTON, ILLINOIS

1979

Abstract

The Hand Test (Wagner, 1962) was administered to 50 institutionalized elderly subjects. There were 36 female and 14 male subjects with a mean age of 76.94 years and a standard deviation of 10.48 years. A list of possible subjects was initially obtained from the activity directors at each nursing home. They were instructed to list the individuals which were "alert" and would be cooperative. The subjects were given a brief organic screening device to eliminate those subjects of severe or moderate mental impairment. Those subjects which missed two or less questions were then administered the Hand Test.

Following approximately a 35 day interval ($M = 34.90$, $SD = .30$), subjects were again administered the Hand Test. The subjects were unaware that they would be administered the Hand Test twice. It was hypothesized the Hand Test variables would be significantly correlated between administrations. Results indicated that of the 24 investigated variables, 23 were significantly correlated between test administrations. These results can be attributed to the Hand Test's standardized instructions for administration and scoring, it's less ambiguous and complex stimulus cards, and its relatively short length. Another possible reason for the relatively high reliabilities is that the Hand Test purportedly measures aspects of the subjects personality which are closer to

the surface, rather than an in-depth look, as reflected by the Rorschach, Thematic Apperception Test, and the Holtzman Inkblot Technique. The results can also be attributed to the short, rigid, and stereotypical responses given by the institutionalized elderly. The results also suggest the elderly maintain a stable personality structure, as measured by the Hand Test, over a 35 day interval. Implications of the findings are discussed and suggestions for future research in the area were made.

TABLE OF CONTENTS

	<u>Page</u>
List of tables.....	ii
Acknowledgements.....	iii
Introduction.....	1
Review of literature.....	6
Reliabilities of the Rorschach.....	6
Reliabilities of the Holtzman Inkblot Technique...	12
Reliabilities of the Thematic Apperception Test...	18
Reliabilities of the Hand Test.....	23
Hypothesis.....	25
Method.....	26
Subjects.....	26
Materials.....	26
Procedure.....	27
Results.....	28
Table 1.....	29
Table 2.....	30
Discussion.....	31
Footnotes.....	34
References.....	35
Appendix A.....	44

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1 Test-retest correlations for the Hand Test variables.....	29
2 Means, medians, and standard deviations for the Hand Test variables.....	30

ACKNOWLEDGEMENTS

I wish to sincerely thank my chairperson, Dr. Paul Panek, for his time and patience in the supervision of this thesis. I want to also thank Sue Stoner and Dr. James Kantner for serving on my thesis committee and providing support when I needed it most.

I would also like to express my appreciation and love to my wife Judy. She made the most sacrifices of anyone toward the completion of this paper.

Chapter I
Introduction

The demonstration of sufficient reliability has been a problem for all psychometric tests, particularly with projective techniques (Zubin, Eron, & Schumer, 1965). Some of the problems peculiar to projective techniques have been explored by Murstein (1963). He pointed out six basic problems:

1. Projective techniques are less standardized in administration and scoring than more objective devices (e.g. intelligence tests).
2. The subject may respond to any part of the stimulus he wishes.
3. The ability of the scorer may have a significant effect on the reliability of the instrument.
4. The scoring philosophy of the examiner may have significant effects on its interpretation (i.e. descriptive content categories, formalistic categories, interpretive categories, etc.).
5. The physical presence of the examiner has differing effects on different subjects.
6. The length of the tests and the verbal fluency of the subject effects reliability. (Longer tests are commonly believed to possess less reliability than shorter devices. Verbally fluent persons give more complex and lengthy responses than verbally depressed subjects.)

It can be noted that the six basic problems listed above are not interrelated and are therefore applicable to a number of reliability measures. The most widely used forms of reliability estimate, as presented by Holtzman, Thorpe, Swartz and Herron (1961) are:

1. Intra-scorer reliability is defined as the degree of agreement obtained when the same protocols have been scored on two different occasions by the individual.
2. Inter-scorer reliability is defined as the degree of scoring agreement between two independent scorers.
3. Test-retest is defined as the agreement between scores obtained by testing the same individuals twice with a fixed interval between tests.
4. The split-half reliability is the degree of agreement between two equal halves of the same instrument.
5. Alternate-form reliability is the degree of agreement between the original instrument and another form which is believed to be equal.
6. Kuder-Richardson reliability is a mathematical technique for estimating the inter-item consistency. This form is used when one assumes that all items are equivalent, which typically can't be assumed with projective techniques.

As can be seen, different forms of reliability measure different forms of agreement or stability, and each has its difficulties, particularly with projective techniques.

There are test users and researchers that are not concerned with the reliability of the instrument used. Cronbach and Gleser (1965) insist that projective tests are instruments that provide a higher breadth of coverage at the expense of lowered fidelity or dependability of information. An earlier supporter of this view was Holzberg (1947) who stated that the concept of reliability loses importance when the purpose of the technique of personality assessment is not so much to measure as it is to describe it.

Not only has reliability in general been questioned for projective techniques, but specific forms of reliability in regards to specific tests. Responding to a clinician's questions of reliability Murstein (1963) attacked this blanket disregard for reliability measures. He stressed the importance of having a significant scorer reliability because, with projective techniques, having low scorer reliability can be an imposing handicap to clinical use. If two scorers cannot agree on how to score a protocol, it is foolish to concern oneself with any other forms of reliability.

Another reliability measure has been questioned, and particularly with the Rorschach, is test-retest reliability. Anastasi (1976) suggests it is questionable because of the personality changes that occur during the interval, and the effects of memory. Also, the subject may be responding to different cues of the stimulus cards. Kenny and Bijou (1953) believe the only way to deal with the memory factor

in test-retest reliability, particularly with the Thematic Apperception Test, is to instruct the subject to give a different story. They conclude their study by stating that it is probably impossible to estimate the test-retest reliability in the psychometric sense.

Rickers-Ovsiankina (1977) stated that the degree of reliability coefficient required, and the type, is a function of the desired use of the instrument. If the test is being used as a gross or wide band estimate of personality, then reliability is not of great concern. She also reports that without adequate interscorer reliability, as did Murstein (1963), any use of the instrument is suspect. Gulliksen (1950) suggested that reliability coefficients be at least .90, and a coefficient below .80 leaves a test suspect.

Despite the controversy and criticisms concerning types, degrees of, and applicability of reliability data with projective techniques, there has been a plethora of published reliability studies with projectives.

In the past several years the Hand Test has been the subject of much research. As with most techniques, their reliability and validity must be established before other research can be conducted. Campos (1968) and Wagner (1978) state there have been few reliability studies of the Hand Test and only one study of test-retest reliability and none with clinical populations. Test-retest reliability must be established if the Hand Test is to be used as a pre

and post treatment measure of success of psychotherapy
or personality change with age.

Chapter II

Review of Literature

This review explores the different forms of reliability measure as applied to several projective techniques. The techniques considered similar to the Hand Test are those in which a stimulus card is presented and the subject is required to verbally respond to a card. The present study is concerned with reliabilities of different populations, specifically adults, the aged, and the test-retest reliability of the Hand Test (HT) (Wagner, 1962a). The major research question of the literature review is what are the reliabilities of the projective techniques similar to the Hand Test, with emphasis on the test-retest reliability.

Reliabilities of the Rorschach

The interscorer reliability of the Rorschach has been investigated intensively. Levine and Spivak (1964) reviewed four such studies and reported a correlation of .95 or better. Perhaps one of the first interscorer reliability studies on the Rorschach, Vernon (1935) obtained a 93% agreement between scorers. However, it should be noted that there was a prior agreement on scoring between the two raters. A study by Sicha and Sicha (1936) used 5 investigators who weren't allowed to discuss scoring and 300 Rorschach responses were scored.

The investigators reported a 70-80% agreement between scorers. Ramzy and Pickard (1949) examined the interscorer reliability using 50 Rorschach protocols consisting of 673 responses. These experienced scorers reported a 99% agreement. Rieman (1953) obtained a coefficient of .92 using the content categories of the Rorschach. By using the scoring of populars to test interscorer reliability, he reported an agreement of 82%. The populars were scored using the Hertz criterion.

Using 8 graduate students, Elizur (1949), reported a mean reliability of .77. The scorers independently scored 15 Rorschach protocols. The mean correlation between the average of the eight students scoring and Elizur's own scoring of the protocols was determined to be .89. Forsyth (1959) later conducted a study using Elizur's content system. He reported interscorer reliabilities of .90 and .95.

Interscorer reliability studies have been conducted using only the Barrier and Penetration scales of the Rorschach. Fisher and Cleveland (1958) scored 20 protocols for Barrier and Penetration scores which resulted in correlations of .82 and .94 respectively. Again in 1968, they repeated their study using 20 records scored by themselves. In this later study they obtained higher coefficients: .97 for Barrier and .99 for Penetration. It was their conclusion from these

studies that well trained and motivated judges can generally agree somewhere in the .90 range.

Their conclusion was supported by a study conducted by Ramer (1963) in which 3 trained judges scored 96 female undergraduate protocols for Barrier scores. Their mean correlation was .89. Goldfield, Stricker, and Wiener (1971) investigated these same scales for interscorer reliability. Their supporting coefficients were also above the .90 level. Fisher and Cleveland (1968) also studied these scales and their coefficients were consistently over .90. Eigenbroder and Shipman (1960) also used the Barrier scale, however, their correlation was .80. Gulliksen (1950), however, reported coefficients over .80 to be acceptable. Therefore, one could conclude, on the basis of the above investigations, that the Rorschach demonstrates adequate interscorer reliability.

The test-retest form of reliability of the Rorschach has been studied extensively. In a review of these studies using the simple content scales, Aronou and Reznikoff (1976) state the results of these studies to be inconsistent and in general disappointing.

Kagan (1960) administered the Rorschach to 37 male and 38 female subjects with a retest interval of three years. He scored these protocols for 2 categories and the resulting coefficients were all below statistical significance. The range of reliability coefficients were

from .00 to .25.

Eichler (1951) administered the Rorschach to 35 male college students with a median interval of 21 days. The resulting coefficients for the scored categories H, A and At were .76, .74, and .76 respectively. Holzberg and Wexler (1950) using 20 chronic schizophrenics as subjects and a three week retest interval obtained very similar results.

A more recent study of the test-retest reliability of the Rorschach using a relatively long retest interval was conducted by Wagner and Daubney (1976). They scored pairs of protocols for each of the 25 neurologically impaired patients with an average interval of three years. They reported significant correlations for all except a few major summary scores. They also reported an additional analysis which revealed no strong relationship between the varying test-retest interval lengths and the stability of various Rorschach scores for the neurologically impaired.

An area of concern with test-retest and the Rorschach is that many of the lower coefficients reported in the literature were when the retest interval was relatively longer and when the population consisted of children and adolescents. It has been argued the differences in the Rorschach categories over time might reflect developmental changes in the subject (Anastasi, 1976). This view receives support from Holzberg (1947) who stated the

test-retest reliability is affected by the length of time between tests and also the age of the subjects. He also said that younger subject's personalities change more between tests than do older subjects.

Another form of reliability used to assess projective devices is the alternate-form type. The main question raised when using this form of reliability is whether the alternate form used is truly equivalent (Holzberg, 1947). Of the five studies examined using the Behn-Rorschach as the alternate form, four reported substantial agreement between forms (Singer, 1952; McFarland, 1954; Buckle & Holt, 1951; Eichler, 1951). The only study which questioned the use of the Behn-Rorschach as an equivalent form was Swift (1944), who tested pre-school children. It could be questioned whether the choice of subjects influenced the reliability coefficients.

Another type of reliability used with projectives as well is the split-half method. An early review of the literature conducted by Thornton and Guilford (1936) reported contradictory findings of split-half reliability for the Rorschach. An examination of the literature supports their inconclusive report. Vernon (1933) reported low reliability for all variables except the number of responses. Hertz (1934), Ford (1946), Wirt and McReynolds (1953), and Zubin, Eron, and Sultan (1956)

all reported obtaining reliability coefficients ranging from .59 to .97, with a median of .70 which is below that believed acceptable by Gulliksen (1950).

The split-half form of reliability is reported by Rosenzweig, Ludwig and Adelman (1975) to be inappropriate for the Rorschach because the Rorschach consists of an odd number of colored cards and an odd number of black and white cards. They also contend that each blot is unique and therefore equal halves are impossible to construct. This concern might also be considered true for other projective devices, as well as the Hand Test.

A type of reliability which seems similar to test-retest and alternate-form reliability is the delayed alternate-form type of reliability. Two studies using this forms of reliability with the Rorschach were conducted by Swift (1944) and Eichler (1951) who reported contradictory findings. Both studies used the Behn-Rorschach as the alternate form. Swift (1944) used a one week delayed interval and reported a mean reliability coefficient of .74, while Eichler (1951) had a median interval of three weeks. He reported a median reliability coefficient of .55. It appears these studies suggest that the seemingly simultaneous measure of two reliabilities, which appears in the delayed alternate form, yields data which may be difficult to interpret.

In a critique of the studies concerning the Rorschach, Arnou and Reznikoff (1976) stated the lower coefficients

of test-retest reliability were obtained when the interval was longer and when the population was composed of young subjects. It has been argued that differences in the response categories over time might reflect changes in the subject (Anastasi, 1976).

Symonds (1949) states that the concept of reliability loses importance when the purpose of the technique of personality assessment is not so much to measure as it is to describe it. This belief is supported by Holzberg (1977) who states that the problems with assessing the Rorschach is inappropriate for the purpose for which the Rorschach is being used in clinical practice. Vernon (1935) described the Rorschach as analogous to a play technique in that it is not a test in the usual sense of the word, but a means of obtaining insight into the personality. Therefore, the studies of reliabilities of the Rorschach yield conflicting, if not questionable data.

Reliabilities of the Holtzman Inkblot Technique

Unlike most projectives, the Holtzman Inkblot Technique (HIT) is the result of an attempt to eliminate the technical deficiencies found in the Rorschach (Holtzman, Thorpe, & Schwartz, 1961). The HIT consists of two series of 45 cards each and the subject's responses are held to one per card. By doing so, many of the pitfalls of reliability measures with the Rorschach are avoided (Holtzman, et al., 1961).

Holtzman examined the interscorer reliability of the HIT using 40 schizophrenics as subjects and two highly trained scorers (1961). Using the Human, Animal, and Anatomy categories to determine the interscorer reliability, the researchers found coefficients of .99 for Human responses, .99 for Animal responses, and .98 for Anatomy responses. Holtzman, et al. (1961) also examined the interscorer reliability using the Barrier and Penetration scales. They used the same subjects and scorers as reported in the previous study. The correlations for the Barrier and Penetration scales were .95 and .92 respectively. In a later study by Megargee (1965) he, and another scorer, examined 75 protocols for Barrier scores and obtained an adequate correlation coefficient of .86.

In another study by Holtzman, et al. (1961) interscorer reliability was again examined. In this study six scoring categories were used in examining the protocols of 92 superior college men. The amount of agreement ranged from .91 to .99 with a median correlation of .98. Interscorer reliability was again examined, however, in this later study, 4 scorers with a wide range of experience were used. One was highly trained, 2 had less experience, and one had no experience. They scored 96 protocols from college men using a wide range of scoring criterion. Each examiner independently scored one third of the 24 protocols that had been twice scored by the other examiners. The

coefficients of reliability ranged from .56 to .94 with a median coefficient of .86. On the basis of the above studies, it appears that adequate interscorer reliability has been demonstrated for the HIT.

Intra-scorer reliability was also studied by Holtzman, et al. (1961). They had three examiners rescore 24 protocols each after an interval of several months. They scored thousands of responses in the interim, which makes it less likely to recall their scoring of the test protocols. These 72 test protocols were randomly selected from 120 Texas college students. The three examiners differed in experience, one was highly experienced and the others had relatively little experience. Nine particular scores were chosen because they were the only ones which revealed any significant difference due to examiner error in the factorial design. This scoring reliability was conducted especially to determine the extent to which the obtained examiner differences could be accounted for in terms of scoring bias, rather than real differences in inkblot protocols. The median for the experienced examiner was .99 and for the less experienced examiners, .87 and .84 respectively. These results suggest that the Hit does indeed possess good intra-scorer reliability.

Holtzman, et al. (1961) also examined the test-retest reliability of the HIT. Using 120 college students as

subjects and one week as a retest interval, resulting coefficients ranged from .24 to .69. Using 72 11th graders, with an interval of three months, they obtained coefficients from .25 to .60. When they used 42 elementary school students and a one year retest interval, their results ranged from .11 to .64. Again using a one year interval, but this time with a population of 48 college students, their results were nearly identical with a range of .15 to .64. These results seem to indicate that the HIT is not free of the difficulty with test-retest reliability as did the Rorschach.

Holtzman, et al. (1961) state that, because the HIT has two forms it is possible to estimate the amount of error variance attributable to the combined effects of temporal fluctuations and content sampling. They also state that split-half reliability is usually thought of as characteristic of the test scores alone, rather than jointly of the test, the method of administration, and the population tested. Therefore, greater emphasis on the degree of stability of the subject's responses as a variable is justified. With the HIT, split-half consists of 22 odd and 22 even numbered cards. The 45th card is dropped to ensure equality. In a large study by Holtzman, et al. (1961), 15 different populations were used, as were 21 different test variables. The number of subjects per population ranged from 41 to 197, with a median number of 76. The resultant median correlations

for the 15 populations ranged from 65 to 91. The results of this study suggest that the two halves are statistically significantly similar. However, split-half reliability coefficients tend to be spuriously high as estimates of intra-subject stability because significant factors contributing to error variance are excluded. Such contributing transient factors as the subjects temporary mood, the response set induced by the examiner, and the motivation of the subject are ignored as sources of variance since they affect both halves equally. The above are stable over several days and are of greater interest than the momentary fluctuations. Consequently, split-half coefficients may be considered as an indication of the upper limits of intra-subject stability. The most accurate estimate of intra-subject stability is provided by the intra-class correlation. This correlation is equivalent to the test-retest reliability after elimination of systematic components of variance due to such factors as order of presentation, minor differences in the two test forms, and practice effects over the two trials (Holtzman, et al., 1961). Therefore, they conclude that studies of delayed alternate-form reliability should be conducted with the HIT.

One such study reported by the authors resulted in a correlation between the two forms of between .60 and .76 with a median of .71. The subjects used were 98 college students and the time interval was one year. Three similar studies were also reported by the authors. One

year later 120 college students were randomly assigned to each of the four different authors. Each subject was tested twice with a one week interval between testings. One half was given form A before B and the other half reversed. The results reported were that for only 2 of the 23 variables was there a difference significant at the .01 level. The median correlation was .58.

In a similar study by these same researchers, 48 college students were given the alternate form after an interim of one year. Again the results indicated that all but one variable was significant at the .01 level.

In a discussion of the results of the above studies, Holtzman, et al. (1961) state that all of their test-retest studies demonstrate the closeness of the parallel forms A and B, as well as the general intra-subject stability of the major inkblot variables. The authors also report that only a small number of variables change appreciably with time as far as groups are concerned. They say that intra-class correlations can be too high as well as too low, when one is concerned with the study of individual differences through time. If the correlations are very high, the techniques may be insensitive to normal variations expected when dealing with personality variables. If they are very low, the technique and the traits measured by it are too unstable for most purposes. They state the majority of the correlations they published are moderately high. Therefore, they say this is ample

justification for using the HIT to study changes in perception and personality over a period of many months.

One advantage to the HIT over the Rorschach which could explain the higher correlation coefficients reported for the HIT is that the HIT has greater standardization of scoring and administration than the Rorschach.

Therefore, less variance.

Several studies conducted after Holtzman, et al. (1961) reported the above findings, tend to support Holtzman's findings of alternate form reliability. Fisher and Renik (1966) using 20 female subjects and a short interval obtained a correlation coefficient of .85. Renik and Fisher (1968) replicated the previous experiment and obtained a correlation coefficient of .87.

Reliabilities of the Thematic Apperception Test

The Thematic Apperception Test (TAT) is slightly different from the Rorschach and HIT in both administration and scoring, and reliability coefficients. The 19 TAT cards contain vague black and white pictures which are more highly structured and require more complex responses than inkblots. The TAT is widely used in both practice and research, and has been a model for other similar instruments (Anastasi, 1968).

As has other instruments, the TAT's inter-scorer reliability has been thoroughly examined with the same inconclusive results found in most projective techniques.

Gurin, Veroff, and Feld (1959) used 9 scorers to test

inter-scorer reliability for the TAT. They each scored 2 stories of each subject for only motive, which totaled 3200 scored stories each. The results were that the average score reliability was .77. They concluded from their results that although they had a slight decrease in scoring reliability for the usual coefficients, its value is sufficiently high to guarantee usable data for research purposes.

In a unique study by Davenport (1952), he used 4 psychiatric patients and 2 normal persons as subjects. Their protocols were interpreted separately by six clinicians and their analysis subdivided into 207 interpretive statements. Once a week for six weeks new clinicians saw one of the original records and decided which of the statements applied to the record. The criterion of reliability was agreement by all six judges on whether a statement was applicable or not for at least two of the six patients. Only two of the 207 statements could meet this requirement.

Sanford (1943) scoring the entire series of Murray needs and presses for the TAT protocols, reported average inter-scorer correlations of .57 for needs and .54 for presses.

Feld and Smith (1958) reviewed the inter-scorer reliability in 14 studies employing the McClelland system for scoring n-Ach, n-Affiliation and n-Power. The reliabilities range from .66 to .99 with a median of .89. Even with novice scorers (12 hours training) they obtained

a median reliability of .87.

As noted in the above studies, as the number of scoring categories increases, the correlation coefficient decreases. The only adequate correlation was reported in a study using only three variables.

The agreement of researchers on test-retest reliability is also poor as with other projective devices. Lesser (1961) says that most of the evidence supports the general conclusion that the test-retest reliability of the TAT measures is very low, far below levels of reliability demanded by constructors of tests of mental ability, social abilities, etc..

A study by Kagan (1959) revealed that only two of eight variables showed statistically significant test-retest reliabilities over the course of a six year period. He stated that these two relatively stable variables were elicited primarily by pictures which unambiguously portrayed these variables.

In a review of studies of test-retest reliability of the TAT, Morgan (1953), supported the statement later made by Kagan (1959), by stating the studies he reviewed reported low but significant test-retest reliability data for achievement measures when the stimulus (card) used was relatively clear for the variable.

Lesser (1961) stated that he believes we don't have any evidence of test-retest reliability, and what we do have is longer-interval evidence and shows extremely low

reliability.

Kenny and Bijou (1953), state they think the only way you could get test-retest reliability is to instruct the subjects to give a different story. They are assuming that studies typically measure the amount the original story was recalled. They further report that it is probably impossible to estimate the test-retest reliability in the psychometric sense.

Lindzey and Herman (1955) attempted to examine the test-retest reliability of the TAT and keep Kenny's statements in mind. They instructed 20 subjects, 10 highly prejudiced and 10 non-prejudiced, not to repeat their earlier stories. They examined variables and after a two month interval only three variables had a correlation significantly above zero.

A group of Navy men were retested after a month at sea. The correlation coefficients for the variables examined was .13 (Auld, Eron, & Laffal, 1955). They further conclude that overall, there appears to be a very low but significant correlation between test and retest. His conclusion, based on a correlation coefficient of .13, is questionable.

Kagan (1959) in an attempt to explain the high degree of variability in test-retest reliability studies of the TAT, said that the presence of high stimulus structure enhances reliability. Haber and Alpert's study (1958) supported this conclusion. They obtained a test-retest reliability of .36 for a low n-Ach stimulus structure

cards and .59 for high cue cards after a three week period.

Auld, et al. (1955) also state in their study that the maximum decrease in test-retest reliability occurs during the first two months. An earlier study by Tomkins (1947) supports Auld, et al.'s. conclusion. They studied three groups of 15 women, each at intervals of two, six, and ten months. Using Murray's need-press method of scoring the TAT protocols, the correlation coefficients were .80 at two months, .60 at six months and .50 at ten months. However, the studies previously examined failed to support their hypothesis. Several studies had intervals of one and two months, yet failed to obtain adequate reliability coefficients.

In the study of the internal consistency of the TAT, Auld, et al. (1955) used Kuder-Richardson formula 14 and obtained a reliability of only .43 for a gross measure, such as whether or not the story had any sexual contents.

Murstein (1963) reviewed studies of internal consistency of the TAT and said the internal consistency values for the studies reported are, generally speaking, quite low. They indicate that the portions of the tests compared rarely manifest equal representation of the need they studied. This means that much of the response can be attributed to the stimulus-pull of the card, and unless the cards are selected on basis of scaled values, it is unrealistic to expect high internal consistencies (Murstein, 1963). He further states that it is impossible

to obtain an accurate estimate of the reliability of a projective technique. Further, test-retest reliability may be low because of mood fluctuations and personality changes between tests. Two equal halves are required for a split-half reliability test, which the TAT and the Rorschach don't have. Plus, parallel forms are not feasible with projective techniques because of the complex stimuli involved.

Reliabilities of the Hand Test

The Hand Test (HT) (Wagner, 1962a) has been used with various populations (Wagner & Capatasto, 1966; Wagner, 1962b; Levine & Wagner, 1974) and has current norms for the aged (Panek, Wagner & Avolio, 1978). However, as with other projective techniques, determination of adequate reliability has been a problem, especially with test-retest reliability, which has yet to be adequately demonstrated.

The HT escapes some of the previous criticisms because of its standardized instructions for administration and scoring, and because of the less ambiguous and complex stimulus cards. The relatively short length of the HT also contributes to its demonstrated reliability (Wagner, 1962a).

Wagner (1962a) using 100 randomly selected protocols, attempted to establish inter-scorer and split half reliabilities. Three novice scorers were used with only the manual as an aid. Using Pathology score (PATH) as the test variable, the PATH coefficients for the three

scorers, using odd-even cards for split-half reliability were, $A = .85$, $B = .84$, and $C = .88$. Therefore, adequate split-half reliability was established in the study.

To establish the inter-scorer reliability, Wagner (1962a) defined agreement as perfect identity on any of the 15 specific scoring categories, for each of the responses, on all of the 100 protocols. Errors of both commission and omission were counted as mistakes. The inter-scorer reliabilities of the same three scorers are $A + B = .78$ and $B + C = .83$. Therefore, Wagner has demonstrated adequate inter-scorer reliability using three naive scorers.

Campos (1968) stated that it would seem that further reliability data are needed, particularly test-retest or stability indices. And, as a new test, major research applications of the HT have been addressed to its validation. Wagner (1978) states there have been few reliability studies. Those that have been done have been split-half and inter-scorer reliabilities (Wagner, 1962a). There has been one study using test-retest reliability (Panek & Stoner, 1979) using college students as subjects.

On the basis of this literature review, the present study proposes to further explore the test-retest reliability of the HT variables using institutionalized elderly subjects. As can be noted by the literature, no reliability studies have been done using the elderly, although the Hand Test has and continues to be used with this population (Panek, Sterns & Wagner, 1976; Panek & Rush, 1979).

Hypothesis: The correlation between the test and retest administrations for all of the Hand Test variables will be significant.

This hypothesis is based on the studies of the institutionalized elderly by Webb (1959) and Fogel, Swepston, Zintek, Vernier, Fitzgerald, Marnocha & Weschler (1956) whose results indicate these subjects gave more rigid, stereotypical, withdrawn, less creative, and therefore, more restricted responses than noninstitutionalized subjects. These responses would be conducive to higher test-retest reliabilities. Ames (1954, 1960, 1968, 1973) supported this by stating that older subjects give more restricted responses.

Chapter III

Method

Subjects

The subjects were 50 elderly volunteers from the central Illinois area. Five nursing homes were contacted and of a combined total population of 507, only 50 subjects were appropriate and available for retesting. No subject tested was bed-ridden, they were all partially or totally ambulatory, and without severe sensory or motor impairment. Subjects were 14 male and 36 females. The median age was 78.50 years, the mean age was 76.94 years, and the standard deviation was 10.48 years.

Subjects were not chosen a priori by sex, however, as noted by Rockstein (1958), females live longer and maintain better physical condition later in life than males. Also, Rhudick and Gordon (1973), and Jarvik, Eisdorfer, and Blum (1973), suggest that the intellectual functioning of females decline less rapidly than do males. Therefore, more females were appropriate as subjects than males.

Materials

The Hand Test consists of 10 cards, nine depicting hands in various positions and one blank. The subjects are asked to explain what the hand is doing while the examiner records their responses verbatim (Wagner, 1962a).

The Mental Status Questionnaire (Kahn, Pollack, & Goldfarb, 1973) consists of 10 questions which the examiner asks the subject and records their responses.

The content of the Questionnaire (see Appendix A) consists of questions pertaining to awareness of self, immediate surroundings, and gross ideas of current events.

Procedure

All subjects were obtained by the activity directors of each nursing home. They were instructed to list the individuals which were "alert" and would be cooperative. All subjects were then administered the Mental Status Questionnaire in order to avoid testing those subjects of severe or moderate mental impairment (Kahn, et al., 1973). Those subjects which missed two or less questions were then administered the Hand Test. Following a mean interval of 34.90 days, the subjects were again administered the Hand Test. The standard deviation for the retest interval was .30 days. All subjects were unaware that they would be administered the Hand Test twice. All Hand Test protocols were scored blind by an expert¹.

Chapter IV

Results

Pearson-product moment correlations (r) were computed between the test and retest administrations for each of the Hand Test variables. These correlations are presented in Table 1. Twenty-three significant correlations were found.

In addition, split-half (i.e., odd²-even³) reliabilities were calculated for each protocol, for each test administration and for the variables that go into the calculation of the Pathology Score (i.e., Tension, Crippled, Fear, Description, Bizarre, Failure). The correlation for the number of pathological responses for the even numbered cards between administrations was ($r=.52$, $p < .001$). The correlation for the number of pathological responses for the odd numbered cards between test administrations was ($r=.59$, $p < .001$). The odd-even correlations for the number of pathological responses with the overall Pathology score in the first administration was ($r=.85$, $p < .001$), and ($r=.91$, $p < .001$), respectively. For the second administration the odd-even correlations for the number of pathological responses with the overall Pathology Score were ($r=.78$, $p < .001$), and ($r=.86$, $p < .001$), respectively.

The Means, Medians, and Standard Deviations for the Hand Test variables for both the test and retest administrations are presented in Table 2.

Table 1

Test-retest correlations for the Hand Test variables ($n = 50$)

<u>Variable</u>	<u>r</u>	<u>Variable</u>	<u>r</u>
Affection	.67***	Crippled	.47***
Dependence	.69***	Fear	.12
Communication	.70***	Maladjustive	.47***
Exhibition	.52***	Description	.73***
Direction	.69***	Bizarre	.63***
Aggression	.41***	Failure	.61***
Interpersonal	.83***	Withdrawal	.71***
Acquisition	.69***	Number of Responses	.71***
Active	.40***	Average Initial Reaction Time	.44***
Passive	.43***	High minus Low	.29**
Environmental	.58***	Pathology	.71***
Tension	.44***	Acting Out Score ^a	.44***

** $p < .01$.

*** $p < .001$.

^anote: Acting Out Score = (Direction + Aggression) -
(Affection + Dependence + Communication).

Table 2

Means, Medians, and Standard Deviations for the Hand Test Variables ($n = 50$)

Variables	Test			Re-test		
	\bar{X}	Mdn	SD	\bar{X}	Mdn	SD
Affection	.82	.73	.85	.92	.58	1.12
Dependence	.56	.33	.81	.66	.33	.96
Communication	.98	.50	1.35	.80	.43	1.11
Exhibition	.24	.11	.59	.18	.11	.39
Direction	.84	.46	1.04	.70	.50	.84
Aggression	.64	.50	.75	.64	.59	.69
Interpersonal	4.10	3.83	3.11	3.90	3.90	2.82
Acquisition	.34	.14	.75	.32	.18	.62
Active	3.34	2.96	1.95	3.30	3.13	1.88
Passive	.44	.24	.67	.46	.31	.65
Environmental	4.10	3.77	2.20	4.08	3.89	2.17
Tension	.28	.13	.64	.38	.21	.64
Crippled	.42	.24	.73	.50	.24	.86
Fear	.12	.07	.33	.08	.04	.27
Maladjustive	.82	.43	1.16	.96	.50	1.20
Description	1.36	.50	2.05	1.54	.68	2.25
Failure	.82	.36	1.34	.56	.36	.81
Bizarre	.10	.04	.36	.06	.03	.24
Withdrawal	2.28	1.70	2.41	2.16	1.43	2.32
Responses	10.50	10.28	2.94	10.54	10.13	2.13
Average Initial	4.17	3.67	4.63	5.03	3.56	4.03
Reaction Time						
High Minus Low	.13	.09	.16	.13	.08	.14
Pathology	5.38	4.30	4.77	5.26	4.00	4.59
Acting Out	.88	.77	2.01	1.04	.72	2.00
Score						

Chapter V
Discussion

Projective test responses are typically thought to be effected by several problems, such as, lack of standardization in administration and scoring, the length of the test, and the ambiguous and complex nature of the stimulus cards (Murstein, 1963). However, the results of the present investigation indicate that the Hand Test responses remain reasonably consistent within an individual over a five-week interval. Perhaps one reason for the test-retest reliabilities to be higher than most other projective techniques pertains to its standardized instructions for administration and scoring, its less ambiguous and complex stimulus cards, and its relatively short length. Another possible reason for these relatively high reliabilities is that the Hand Test purportedly measures aspects of the subjects personality which are closer to the surface (Wagner, 1962a), rather than an in-depth look, as reflected by the Rorschach, Holtzman Inkblot Technique, and the Thematic Apperception Test. These results might also suggest the aged subject maintains a relatively more stable personality, as measured by the Hand Test, than younger subjects over a 35 day interval.

The elderly typically give more rigid, stereotypical, withdrawn, less creative and more restricted responses than younger and noninstitutionalized subjects (Ames,

1960; Fogel, Swepston, Zintek, Vernier, Fitzgerald, Marnocha, & Weschler, 1956; Webb, 1959). This previous research on the responses of the aged is supported by the results of this study. The mean number of responses for the test and retest are 10.50 and 10.54 respectively. However, in the test-retest investigation by Panek and Stoner (1979), who used young subjects (M age = 19.18 years), the mean number of responses were 14.00 and 14.11 for the test and retest respectively. Therefore, the elderly subjects are more restricted in terms of responses, than younger subjects.

Perhaps, though the correlation coefficients were modestly significant, this significance may be greater than it appears. Anastasi (1968) states that the more homogeneous the sample tested, the smaller the resulting correlations. When this is taken into consideration, the correlations obtained would possibly have been higher had the sample been more heterogeneous and the range of the sample been less restricted.

Though the present study has limitations due to the fact that only institutionalized elderly subjects, primarily females, were used, the findings of the present investigation have limitations for present and future uses of the Hand Test. For example, since the variables were found to be reliable in a test-retest situation, the Hand Test appears to be an adequate instrument for use with

the elderly, specifically for evaluating the effects of psychotherapy or counseling, and evaluating changes in personality with age. Future investigations of the test-retest reliability of the Hand Test could focus on different clinical populations and with differing intervals.

Footnotes

¹All Hand Test protocols were scored blind by Dr. Paul Panek.

²note: Cards I, III, V, VII, IX.

³note: Cards II, IV, VI, VIII, X.

References

- Ames, L. B. Other projective techniques. In: Robin, A. I. (ed.), Projective Techniques in Personality Assessment, a Modern Introduction, New York: Springer, 1968.
- Ames, L. B. Age changes in the Rorschach responses of a group of elderly individuals. The Journal of Genetic Psychology, 1960, 97, 257-285.
- Ames, L. B., Metraux, R. W., Rodell, J. L., & Walker, R. N. Rorschach Responses in Old Age. New York: Brunner-Mazel, 1973.
- Ames, L. B., Metraux, R. W., Rodell, J. L., & Walker, R. N. Rorschach Responses in Old Age. (2nd ed.) New York: Brunner-Mazel, 1973.
- Anastasi, A. Psychological Testing, (4th ed.) New York: Macmillan, 1976.
- Aronow, E., & Reznikoff, M. Rorschach Content Interpretation, New York: Grune & Stratton, 1976.
- Auld, F., Eron, L. D. & Laffal, J. Application of Guttman's scaling method to the TAT. Educational and Psychological Measurement, 1955, 15, 422-435.
- Buckle, D., and Holt, N. Comparison of Rorschach-like projective tests. Journal of Projective Techniques, 1951, 15, 486-493.
- Campos, L. P. Other projective techniques. In: Robin, A. I. (ed.), Projective Techniques in Personality Assessment: A Modern Introduction. New York: Springer, 1968.

- 50
- Cronbach, L. J., & Gleser, G. C. Psychological Tests and Personnel Decisions, (2nd ed.) Urbana, Ill.: University of Illinois Press, 1965.
- Eichler, R. M. A comparison of the Rorschach and Ben-Rorschach inkblot tests. Journal of Consulting Psychology, 1951, 15, 185-189.
- Eigenbrode, C. R., & Shipman, W. G. The body image barrier concept. Journal of Abnormal Psychology, 1960, 60, 450-452.
- Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. Rorschach Research Exchange and Journal of Projective Techniques, 1959, 23, 207-213.
- Fisher, S., & Cleveland, S. E. Body Image and Personality, Princeton, N. J.: Van Nostrand, 1958.
- Fisher, S., & Cleveland, S. E. Body Image and Personality, (2nd ed.) New York: Dover, 1968.
- Fisher, S., & Renik, O. D. Induction of body image boundary changes. Journal of Projective Techniques and Personality Assessment, 1966, 30, 429-434.
- Fogel, E. J., Swepston, E. R., Zintek, S. S., Vernier, C. N., Fitzgerald, J. F., Marnocha, R. S., & Weschler, C. H. Problems of aging; conclusions derived from two years of interdisciplinary study of domiciliary members in a veterans administration center. American Journal of Psychiatry, 1956, 112, 724-730.

- Ford, M. The application of the Rorschach test to young children. University of Minnesota Institute for Child Welfare Monograph, 1946, No. 23.
- Forsyth, R. P. The influence of color, shading and Welsh anxiety level on Elizur Rorschach content test analysis of anxiety and hostility. Journal of Projective Techniques, 1959, 23, 207-213.
- Goldfried, M. R., Stricker, G., & Wiener, I. B. Rorschach Handbook of Clinical Research Applications. Englewood Cliffs, N. J.: Prentice-Hall, 1971.
- Gulliksen, H. Theory of Mental Tests. New York: John Wiley & Sons, 1950.
- Guren, G., Veroff, J., & Feld, S. Americans View their Mental Health: A Nationwide Interview Study. New York: Basic Books, 1959.
- Haber, R. N., & Alpert, R. The role of situation and picture cues in projective measurement of the achievement motive. In: Atkinson, J. W. (ed.), Motives in Fantasy, Action and Society. Princeton, N. J.: Van Nostrand, 1958, pp. 644-663.
- Hertz, M. R. The reliability of the Rorschach inkblot test. Journal of Applied Psychology, 1934, 18, 461-477.
- Holzberg, J. D., & Wexler, M. The predictability of schizophrenic performance on the Rorschach test. Journal of Consulting Psychology, 1950, 14, 395-399.

- Holzberg, J. D. Reliability re-examined. In: Rickers-Ovsiankera, M.A. (ed.) Rorschach Psychology. Huntington, New York: Robert E. Kruger, 1977, 555-574.
- Holtzman, W. H., Thorpe, J. S., Schwartz, J. D., & Herron, E. W. Inkblot Perception and Personality. Austin, Texas: University of Texas Press, 1961.
- Jarvik, L. F., Eisdorfer, C., & Blum, J. E. Intellectual Functioning in Adults, New York: Springer, 1973.
- Kagan, J. The stability of TAT fantasy and stimulus ambiguity. Journal of Consulting Psychology, 1959, 17, 289-292.
- Kagan, J. The long term stability of selected Rorschach responses. Journal of Consulting Psychology, 1960, 24, 67-73.
- Kagan, J., & Lesser, G. S. (eds.) Contemporary Issues in Thematic Apperceptive Methods, Springfield, Ill.: Charles C. Thomas, 1961.
- Kahn, R. L., Pollack, M., & Goldfarb, A. I. Factors related to individual differences in mental status of institutionalized aged. New York State Department of Mental Hygiene, Office for Aging. In: Brantl, V. M., & Brown, M. B. (eds.) Readings in Gerontology, St. Louis: The C. B. Masby Company, 1973.
- Kenny, D. T., & Bijou, S. W. Ambiguity of pictures and extent of personality factors in fantasy responses. Journal of Consulting Psychology, 1953, 17, 238-288.

- Krumholtz, J. D., & Farquhar, W. W. Reliability and validity on the n-Achievement test. Journal of Consulting Psychology, 1957, 21, 226-229.
- Lesser, R. S. Discussion of Dr. Murstein's paper. In: Kagan, J. & Lesser, G. S. (eds.). Contemporary Issues in Thematic Apperceptive Methods, Springfield, Ill.: Charles C. Thomas, 1961, pp. 274-287.
- Levine, E. S., Wagner, E. E. Personality patterns of deaf persons: An interpretation based on research with the Hand Test. Perceptual and Motor Skills, Monograph Supplement, 1974, 39, 1167-1236.
- Levine, M., & Spivack, G. The Rorschach Index of Repressive Style, Springfield, Ill.: Charles C. Thomas, 1964.
- Lindzey, G., & Herman, P. S. Thematic Apperception Test: A note on reliability and situational validity. Journal of Projective Techniques, 1955, 18, 36-42.
- MacFarlane, J. W., & Luddenham, R. D. Problems in the validation of projective techniques. In: Anderson and Anderson (eds.), An Introduction to Projective Techniques and Other Devices for Understanding the Dynamics of Human Behavior, New York: Prentice-Hall, 1951.
- McFarland, R. Perceptual consistency in Rorschach-like projective tests. Journal of Projective Techniques, 1954, 18, 368-378.
- Megargee, E. I. Relation between barrier scores and aggressive behavior. Journal of Abnormal Psychology, 1965, 70, 307-311.

- Morgan, H. H. Measuring achievement motivation with picture interpretations. Journal of Consulting Psychology, 1953, 17, 289-292.
- Murstein, B. I. Theory and Research in Projective Techniques (emphasizing the TAT) New York: John Wiley & Sons, 1963.
- Panek, P. E., Sterns, H. L., & Wagner, E. E. An exploratory investigation of the personality correlates of aging using the Hand Test. Perceptual and Motor Skills, 1976, 43, 331-336.
- Panek, P. E., & Rush, M. C. Intellectual and personality differences between community-living and institutionalized older adult females. Experimental Aging Research, 1979, 5, 239-250.
- Panek, P. E., & Stoner, S. Test-retest reliability of the Hand Test with normal subjects. Journal of Personality Assessment, 1979, 43, 135-137.
- Panek, P. E., Wagner, E. E., & Avolio, B. J. Differences in Hand Test responses of healthy females across the life span. Journal of Personality, 1978, 42, 113-120.
- Ramer, J. The Rorschach barrier score and social behavior. Journal of Consulting Psychology, 1963, 27, 525-531.
- Ramzy, I., & Pickard, P. M. A study in the reliability of scoring the Rorschach inkblot test. Journal of General Psychology, 1949, 40, 3-10.
- Renik, O. D., & Fisher, S. Induction of body image boundary changes. Journal of Projective Techniques and Personality Assessment, 1968, 32, 45-48.

- Rickers-Ovsiankina, M. A. (ed.) Rorschach Psychology,
Huntington, New York: Robert E. Kruger, 1977,
pp. 555-574.
- Rieman, G. W. The effectiveness of Rorschach elements in
thedescrimination between neurotic and ambulatory
schizophrenic subjects. Journal of Consulting
Psychology, 1953, 17, 25-31.
- Rhudick, P. J., & Gordon, C. The age center of New
England study. In: Jarvik, L. F., Eisdorfer, C., &
Blum, J. E. (eds.) Intellectual Functioning in Adults,
New York: Springer, 1973.
- Rockstein, M. Heredity and longevity of male and female
house flies. Journal of Gerontology, 1957, 12,
253-256.
- Rosenzweig, S., Ludwig, D. J., & Adelman, S. Retest
reliability of the Rosenzweig Picture-Frustration study
and similar semi-projective techniques. Journal of
Personality Assessment, 1975, 39, 3-12.
- Sanford, R. N. Physique, personality and scholarship.
Monograph of Social Research and Child Development,
1943, 8, (1).
- Sicha, K., & Sicha, M. A step towards the standardization
of the scoring of the Rorschach test. Rorschach
Research Exchange, 1936, 1, 95-101.
- Singer, J. The Behn-Rorschach inkblots: A preliminary
comparison with the original Rorschach series.
Journal of Projective Techniques, 1952, 16, 238-245.

- 42
- Swift, J. W. Reliability of Rorschach scoring categories with preschool children. Child Development, 1944, 15, 207-216.
- Symonds, M. Adolescent Fantasy, New York: Columbia University Press, 1949.
- Thornton, G. R., & Guilford, J. P. The reliability and meaning of Erlebnistypus Scores in the Rorschach test. Journal of Abnormal and Social Psychology, 1936, 31, 324-330.
- Tomkins, S. S. The Thematic Apperception Test. New York: Grune & Stratton, 1947.
- Vernon, P. E. Recent work on the Rorschach test. Journal of Mental Science, 1935, 81, 1-27.
- Vernon, P. E. The Rorschach inkblot test. II. British Journal of Medical Psychology, 1933, 13, 179-205.
- Wagner, E. E. The Hand Test: Manual for administration, scoring and interpretation. Los Angeles: Western Psychological Services, 1962a.
- Wagner, E. E. The use of drawings of hands as a projective medium for differentiating neurotics and schizophrenics. Journal of Clinical Psychology, 1962b, 18, 208-209.
- Wagner, E. E. Diagnostic application of the Hand Test. In: Walman, B. B. (ed.), Clinical Diagnosis of Mental Disorders, New York: Plenum, 1978.
- Wagner, E. E., & Capostato, M. Discrimination of good and poor retarded workers with the Hand Test. American Journal of Mental Deficiency, 1966, 1, 126-128.

- Webb, M. A. Longitudinal sociopsychological study of a randomly selected group of institutionalized veterans. Journal of the American Geriatrics Society, 1959, 7, 730-740.
- Wirt, R., & McReynolds, P. The reliability of Rorschach number of responses. Journal of Projective Techniques, 1953, 17, 493-494.
- Zubin, J., Eron, L. D., & Sultan, F. A psychometric evaluation of the Rorschach experiment. American Journal of Ortho-Psychiatry, 1956, 26, 773-782.
- Zubin, J., Eron, L. D., & Schumer, F. An Experimental Approach to Projective Techniques, New York: Wiley, 1965.

APPENDIX A

Mental Status Questionnaire

1. Where are we now? (Correct name of place)
2. Where is this place? (Correct city)
3. What is today's date? (Day of month)
4. What month is it?
5. What year is it?
6. How old are you?
7. When is your birthday? (Month)
8. What year were you born?
9. Who is President of the United States?
10. Who was President before him?

Score: _____

0-2 OBS absent or mild

3-8 OBS moderate

9-10 OBS severe