

**This dissertation has been  
microfilmed exactly as received**

**69-9867**

**KLIEGER, Douglas Myles, 1942-  
AN INVESTIGATION OF THE INFLUENCE OF  
RESPONSE SETS ON THE HOLTZMAN PROJECTIVE  
TECHNIQUE.**

**Iowa State University, Ph.D., 1968  
Psychology, clinical**

**University Microfilms, Inc., Ann Arbor, Michigan**

AN INVESTIGATION OF THE INFLUENCE OF RESPONSE  
SETS ON THE HOLTZMAN PROJECTIVE TECHNIQUE

by

Douglas Myles Klieger

A Dissertation Submitted to the  
Graduate Faculty in Partial Fulfillment of  
The Requirements for the Degree of  
DOCTOR OF PHILOSOPHY

Major Subject: Psychology

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University  
Ames, Iowa

1968

## TABLE OF CONTENTS

	Page
INTRODUCTION	1
PROBLEM	16
METHOD	19
PROCEDURE	25
RESULTS	31
DISCUSSION	35
SUMMARY	39
REFERENCES	41
ACKNOWLEDGMENTS	44
APPENDIX A	45
APPENDIX B	46
APPENDIX C	47
APPENDIX D	49
APPENDIX E	53
APPENDIX F	54

## INTRODUCTION

The notion that human personality is open to description and assessment is likely as old as man's first reflections. Personality formulations can be seen in the early Greek notions of sanguine, phlegmatic, melancholic, and chloric types of man. Parallel, but not similar, constructs were being formed within the Jewish rabbinical tradition and were written into the general "literature of wandering" such as the Talmud. Personality measurement first appears in Galton's work on the psychology of individual differences. On Galton, Boring (1950) says:

Here one can see the beginnings of measurement in psychology, measurement on a non-psychophysical basis that is. Though today the use of dynamometer pressures, rate of movement etc. would not be called or used as personality measures, Galton's intent was definitely toward the personality aspects as well as the eugenics-inheritance conception of man. Of course the move toward personality measurement was mostly the work of James McKeen Cattell. Galton's wish was to establish an "Inventory of Human Abilities", while Cattell attempted "discovering the constancy of mental processes, then interdependence, and then variation under different circumstances".

Cattell, who had taken a Ph.D. under Wundt in 1886, was more in Galton's individual difference tradition than the Continental ideas of the general man. His dissertation under Wundt concerned individual differences in reactions, something Wundt did not favor (Watson, 1959). However, though Cattell

never apparently lost his enthusiasm for individual differences, the use of "mental tests" faded for lack of validation. The tests simply did not intercorrelate nor did they correlate with academic success (Watson, 1959). Thus with Titchener (who was never a tester) and Cattell looking elsewhere, it is not surprising that personality testing first faded and then revived with Binet and intelligence testing since these two men were training most of American's psychologists at this time. Boring neatly summarizes the influences of Galton, Cattell, and Binet in saying:

Perhaps it is true that America, while giving homage to Wundt, has overlooked Galton, to whom it owes a greater debt. Americans, never insufficiently self-conscious, have looked to Cattell, Wundt's recalcitrant student who insisted on working on individual differences, as their great pioneer in the tests. Cattell himself recognized Galton's priority and genius, but Galton seemed remote to the American testers. If only he had come with Helmholtz to the Chicago World's Fair in 1893! They were within six months of each other in age; but Galton did not come. Then, shortly after, Galton lost out to Binet in the matter of the kind of tests that were going to bring out the individual psychological resources most useful to a nation. The testers concentrated upon intelligence, Spearman's G, not on assessing the variety of capacities which were listed in Galton's inventory of human abilities.

After this initial founding, personality testing disappeared into limbo, with the aid of Watsonian behaviorism. Under the influence of behaviorism, intelligence became what intelligence tests measure. These and similarly fostered tests all had excellent behavioral method but little validity.

In England the specificity of abilities and the factor analytic technique produced great strides in statistics, but little in personality.

This rather bleak picture changed rapidly during and after W.W.I. The war demanded the means for classifying large numbers of people in a quick and orderly fashion, a demand to which the test is well suited. The creation of the Army Alpha and the Psychoneurotic Inventory by Woodworth (Watson, 1959) are representative examples of early tests. During this period, Watson (1959) estimates that over 500 personality inventories (and questionnaires) were created as well as many new intelligence tests. The writers in the personality measurement field were soon busy discussing the various reliabilities and validities found in their tests in addition to the techniques of construction. During the latter part of this period mention was occasionally made of the possibility of factors other than true response and error being present in the tests. This emphasis began with B. D. Wood's 1926 article, "Studies of Achievement Tests" and Mathews (1929) "The Effect of the Order of Printed Response on a Interest Questionnaire." Until the later thirties, the general conception was that response effects were solely functions of varying personality traits. With Lentz's 1938 study, "Acquiescence as a Factor in the Measurement of Personality," and Lorge's 1937 study, "Gen-like: Halo or Reality?", response sets were being named and investigated as separate

phenomena.

Cronbach's 1946 article entitled "Response Sets and Test Validity," was the first review in the area. More importantly, it defined the topic clearly and enumerated many of the sets still under consideration today. Cronbach again reviewed the area in 1950. In both reviews he defined response sets as any tendency to respond to a personality item on a basis other than content. Very little experimental evidence was given and in 1950 response sets were still variables all workers assumed rather than having been explicitly demonstrated.

The work begun by H. Rorschach in 1910 made many similar assumptions. In projective testing employing inkblots, variables were not assumed to be present; rather, discriminability and reliability were assumed. Though Rorschach's work did not become part of American psychology until Beck's 1930 paper (reported in Beck, 1953), once introduced there appeared a vast number of studies, reflecting great acceptance and popularity among workers in the area of personality. This study will focus on the possible convergence of response sets and the inkblot type of projective test and therefore a more thorough examination of these areas follows.

### Response Sets

#### Social desirability

The first of the response sets to be operationalized and to have its effects demonstrated was social desirability (SD).

In 1953, Edwards defined the SD response as a True (T) response to an objective personality item with a socially desirable scale value and a False (F) response to an item with a socially undesirable (SUD) scale value. It was demonstrated that the probability that an item would be endorsed ( $P(T)$ ) was a direct, increasing, monotonic function of the SD scale value (SDSV) of the item where items were rated for SD on a scale from 1: "Extremely SUD;" through 5: "Neutral;" to 9: "Extremely SD."

Originally, it was assumed that SDVSs for a standard set of items obtained from a given sample of raters would be largely a function of the culture from which the raters came. This has not proved to be the case, however. The product-moment correlation between scale values for comparable (with respect to age and SES variables) but culturally diverse groups is generally about .90. Ratings given by Norwegian students correlated .78 with those of American college students (Løevaas, 1958). Students at the University of Beirut, Lebanon, gave ratings that correlated .86 with American students' ratings (Klett & Yaukey, 1959). Iwawaki and Cowen (1964) found that scale values derived from students at the Japanese Defense Academy correlated .90 with American college students' scale values. Cowen and Frankel (1964) indicate correlations of .95 between French and American students' ratings, and even more startling is the fact that the correlation between French and Japanese students' ratings was .85.



Correlation of SD ratings by diverse groups within our own culture such as: novice nuns, Nisei, high school students, and grade school children; with ratings of college students yield coefficients of .85 or better (Edwards, 1964).

These findings also hold for abnormal groups. Taylor (1959) had patients with diagnosed schizophrenia and control Ss rate the items on the MMPI schizophrenia scale for social desirability. The correlation was .98 between the two sets of ratings. Similar values have been reported by Klett (1957) for neuro-psychiatric patients (.88) and by Cowen, Staiman, and Woltizky (1961) for male schizophrenics (.98). The test-retest reliability coefficients for SDSVs are also unusually high. Edwards and Walsh (1963) found that for a group of 176 items the test-retest correlation was .97 for males and .98 for females over a two-week interval. Male-female (within a culture) correlations as high as .97 are reported (Edwards & Walsh, 1963).

In short, there is wide and general agreement among adult raters as to what constitutes desirable and undesirable responses to objective personality items regardless of differences among cultures and subgroups. As wide spread as SD appears to be, it does not seem unwarranted to think that it might appear in other than objective personality measures.

### Acquiescence

The development of the concept of an "acquiescence" response set has closely paralleled the development of the idea

of the SD response set. There are two distinct definitions of acquiescence. Cronbach (1946) defined it as "the tendency to mark items 'True' rather than 'False'" when guessing (p. 14). Implicit in Couch and Keniston's research (1960) is the definition of acquiescence as the tendency to give a True response to an item regardless of item content. The former definition requires responses to be a function of S and item, while the latter definition requires only the S (Diers, 1964). There have been many attempts to explicate both definitions in terms of scale keying, item SDSVs, and factor structures (Diers, 1964; Couch & Keniston, 1960; Edwards and Walsh, 1964; Peabody, 1961 & 1966). However, a paper by Rorer and Goldberg (1965), points out that the two above definitions and the S's possible factual answer are completely confounded. Rorer and Goldberg (1965, p. 818) state:

The fact that an individual gives a preponderance of "true" responses to the items on a personality inventory, or even to any sub-set of items from an inventory, provides no basis on which to conclude that he "acquiesced". For personality, attitude, and interest inventories, unless the items are stated in more than one way, the content and the keying are inevitably confounded, no matter how much statistical legerdemain is performed upon the results. Previous studies concluding that acquiescence response style is an important variable in determining MMPI responses have all done so on the basis of such content-confounded measures.

To date the only investigations into "acquiescence" that allows for the confounding problem are studies by Peabody (1961, 1966). In these studies Peabody used both the

standard California F scale and a reversed F scale. From this an unconfounded attitude measure of acquiescence was possible. To acquiesce in this case a S had to agree to both the standard item and its reversal.

### Falsification

A response set corresponding to the third orthogonal MMPI factor has been only tenuously identified. It seems to be characterized by the tendency of Ss to falsify their responses (Edwards and Walsh, 1964). Indeed, the MMPI lie scale and Marlowe-Crowne SD scale have among the strongest loadings on this factor (Edwards, Diers and Walker, 1962). These scales have some overlapping items (Crowne-Marlowe, 1964). In short, the third factor is not well defined nor is the response set item content question resolved.

## Projective Tests

### Rorschach technique

The more recent history of projectives tests, specifically "inkblot" techniques, is indeed a vast literature with thousands of studies. Holtzman et al. (1961) write:

During the first twenty-five years following Rorschach's death, the projective technique bearing his name developed rapidly as the method par excellence for assessing the motivation, thought processes, and basic personality structure of the individual. Competing systems of scoring and analysis flourished, both in Europe and America. Active proponents of the Rorschach in the United States, especially Samuel Beck, Marguerite Hertz, and Bruno

Klopfer, gradually attracted a large following from psychiatrists, clinical psychologists, and others concerned mainly with the psycho-diagnosis of the abnormal personality. The main stream of academic psychology looked askance at the Rorschach movement, criticizing its cultist character and lack of scientific discipline.

The War and subsequent mushrooming of graduate training in clinical psychology brought about the fusion of the Rorschach movement and a major segment of academic psychology, a union that was not without painful conflict. Countless studies by graduate students soon piled up a wave of criticism from which the Rorschach movement has yet to recover. Although much of this research was irrelevant or too inadequately conceived to provide an appropriate evaluation of the Rorschach method, an impressive number of carefully designed validity studies yielded negative results. In the wake of these experimental studies came the growing realization that the Rorschach had inherent psychometric weaknesses, which cast considerable doubt on the interpretation of quantitative scores, particularly the determinants and ratio-scores derived from them.

In 1954 Zubin indicated seven major shortcomings of the Rorschach:

(1) failure to provide an objective scoring system free of arbitrary conventions and showing high inter-scorer agreement; (2) lack of satisfactory internal consistency or test-retest reliability; (3) failure to provide cogent evidence for clinical validity; (4) failure of the individual Rorschach scoring categories to relate to diagnosis; (5) lack of prognostic or predictive validity with respect to outcome of treatment or later behavior; (6) inability to differentiate between groups of normal subjects; and (7) failure to find any significant relationships between Rorschach scores and intelligence or creative ability. Though many might disagree with Zubin, the failures are critical if a metric rather than intuitive technique is sought.<sup>1</sup>

---

<sup>1</sup>Paraphrased by Holtzman et al. 1961.

It is pointed out by Holtzman et al. (1961; Holtzman, 1959) that much of the confusion about and discussion of the Rorschach arises out of a lack of distinguishing between the technique as a projective tool and as an objective psychometric test. Few have objected to the use of the Rorschach as a phenomenological information gathering system for the skilled clinician. However, the number of practitioners who can use such a system without concern for classifying and enumerating the S's characteristics is small. Reliability and validity follow and are required when metric properties are in question, and this is the very thing that all systems of scoring and classification since Rorschach have attempted.

In discussing the metric problem of this technique Holtzman et al. (1961) write:

In the standard Rorschach, each of ten inkblots is shown the subject, one at a time, with general instructions to tell what he sees in the blot. The subject is encouraged to give as many or as few responses to each card as he wishes by the examiner, who writes down what he says in shorthand form. After the subject completes his associations to the ten inkblots, the examiner goes over each response with the subject in a formal inquiry designed to learn what determinants led to each response and to obtain any additional responses given spontaneously by the subject. Throughout both the performance and inquiry phase of the examination, the examiner attempts to maintain a high degree of rapport with the subject. The actual number of responses (R) in a typical protocol varies from less than ten to over one hundred. The subject may give a fairly uniform number of responses to each of the ten cards, or he may give no response to one card and eight or ten to another. Because the inquiry is always a highly individual matter, tailored to the

particular subject and his responses, the interactive influence of the examiner upon the subject and vice versa is generally high. Given such freedom of response, performance variation due to examiner, and the scoring difficulties which follow, it is small wonder that quantitative studies of Rorschach scores have run head-on into almost insurmountable obstacles.

### Holtzman technique

The next step forward in the development of inkblot projective technique took place with the development of the Holtzman Inkblot technique. Here the focus was on metric, not on phenomenology:

The fundamental question of how to develop psychometrically sound scoring procedures for responses to inkblots while also preserving the rich qualitative, projective material of the Rorschach was approached from a fresh point of view by starting over with an entirely new set of stimulus materials. It seemed apparent that the major limitations in the Rorschach could best be overcome by developing a new technique using more inkblots with simplified procedures for administration. By permitting a subject only one response per card, the number of inkblots could be greatly increased without appreciably extending the administration time. Moreover, if one were to go to the trouble of constructing new inkblots, he could easily construct two or more parallel forms of the blots rather than just one.

Such a technique would have several advantages over the standard Rorschach: (1) the number of responses per individual would be relatively constant; (2) each response would be given to an independent stimulus, avoiding the weaknesses inherent in the Rorschach where all responses are lumped together regardless of whether they are given to the same or different inkblots; (3) a richer variety of stimuli capable of eliciting much more information than the original 10 Rorschach plates would be obtained by making a fresh start in

the production of stimulus materials, especially in view of recent experimental studies of color, movement, shading, and other factors in inkblot perception; and finally, (4) a parallel form of the inkblots could be constructed easily from item-analysis data in the experimental phases of development, and adequate estimates of reliability could be obtained independently for each major variable.

A large amount of information has been collected on the metric properties of the HIT. Four forms of reliability have been assessed: "(1) intra-scorer consistency - the degree of agreement obtained when the same protocols have been scored on two different occasions by the same individual; (2) inter-scorer consistency - the degree of agreement between two independent scorers; (3) immediate intra-subject stability - the degree of internal consistency of a score; and (4) delayed intra-subject stability - the correlation between the two parallel forms, A and B, with a specified interval ..."  
(Holtzman et al., 1961, p. 104).

For number one above Holtzman reports correlation coefficients ranging from .89 to .97 with numerous scorers. Number two is found to have correlations from .89 to .99 with a median of .98 for highly trained scorers, and .73 to .89 for the lesser trained or those with no training at all. Number three is of course a measure of internal consistency. The split-half reliabilities were calculated for each scored variable across a large number of SES groups. Four was also measured across SES variables with counterbalanced orders of presentation and time delays of up to one

year. Their results of three and four are too lengthy to present here; however, Holtzman et al., (1961) summarize these reliabilities thus:

All four of the test-retest studies demonstrated the close comparability of the two parallel forms, A and B, as well as the general intra-subject stability of the major inkblot variables such as Location, Form Definiteness, Color, Shading, and Movement. It can also be said that only a small number of variables change appreciably with time as far as group means are concerned. 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 technique is insensitive to normal variations to be 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. The majority of correlations reported herein are moderately high. Such results are ample justification for using the Holtzman Inkblot Technique to study changes in perception and personality over a period of many months.

Validity of the HIT has thus far been confined to a limited number of clinical samples and studies of the inter-relationship of HIT scores with other variables. There remains one further development upon which the paper rests. In 1963 Swartz and Holtzman published the first paper on the group administration of the HIT. The savings in time and administration with group methods is obvious, and the HIT is well suited to such group procedures. The limit of a single response per blot fits well into a slide projection format. As a result of several preliminary studies a standard group administration method was adopted (see Appendix D). Using



these instructions in another study Holtzman, Moseley, Reiner, and Abbot (1963) found split-half reliabilities as follows:

The split-half reliability coefficients range from a low of .23 for Space to a high of .91 for Location, with a median coefficient value of .64 (see Table 1). These values are very similar to those found for the individual version based on a comparable sample of college students during the standardization for the individual HIT. Estimates of the split-half reliability for the individual version range from a low of .31 for Space to a high of .94 for Location, with a median coefficient value of .71 (Holtzman, et al. Tables 6-4 to 6-23).

Test-retest reliability was discussed as follows:

When the variables having sharply skewed and truncated distributions are omitted (Rejection, Space, and Anatomy), the test-retest reliability coefficients range from a low of .25 for Penetration and Popular to a high of .65 for Location, with a median coefficient value of .49. Though somewhat lower, these results are similar to the test-retest coefficients obtained after an interval between testing of one week for a similar sample of college students during the standardization of the individual HIT.

Test-retest reliability coefficients can be considered lower bounds for the intra-subject stability, just as the split-half correlations for these same variables on the other sample of college students serve as upper bounds. In the case of Movement, for example, one can safely say that the reliability of measurement is somewhere between .63 and .75 for typical college students when the group method is given and scored in a standard manner.

Specifically comparing the group version with the individual method it is found:

Nine of the variables--Form Definiteness, Movement, Integration, Human, Animal, Hostility, Penetration, Rejection, and Pathognomic Verbalization--failed to reveal any differences

whatsoever that could be attributed to method of administration. Five more--Location, Space, Color, Anxiety, and Popular--revealed only minor differences in mean or standard deviation, differences that can easily be corrected by an appropriate constant if the published norms for the individual version are to be employed with the group method. Slight differences in the pattern of intercorrelations with other variables could be attributed to method in the remaining four variables--Form Appropriateness, Shading, Barrier, and Anatomy. Even in these four variables, however, only a small part of the score variance is specific to the method of administration.

The one further difference between the two forms of administration is recording of responses. To this point Holtzman indicates that the individual record form can be used provided all Ss are furnished with additional standard instructions, outlined in Appendix D.

## PROBLEM

The current development of objective personality measures as related to response sets can be summarized as one of demonstrated existence. Response set artifacts cloud and contaminate the content meaning of many measures. The current status of the HIT indicates an increase in reliability and objectivity over the more traditional Rorschach and allows a suitable method of group administration.

In speaking of objective personality tests in general Holtzman et al. (1961, p. 180) say:

Unfortunately they all have the common characteristic of dealing only with the subject's superficial response to items the content of which is often transparent. Although ingenious methods have been developed to disguise this content and force the subject to make choices which can then be scaled and treated psychometrically, the fundamental, superficial nature of such tests still persists. In fairness to the self-inventory approach, it should be pointed out that scales from these tests usually have fairly high reliability and often correlate with socially observable behavior to a higher degree than any projective technique. Such correlations, however, can frequently be traced directly to the fact that the individual has a conscious self-concept that dominates his test responses and is not unrelated to his social behavior as judged by others.

It is this conscious self-concept and related variables that give rise to this paper. If a person's responses to objective assessment devices and his behavior are related to

his social behavior as judged by others, then it would seem that this could be a general trend and not idiosyncratic to paper-and-pencil measures. Holtzman et al. (1961) do not consider this to be so. On page 181 he states:

Clearly, there is little relationship between personality traits measured by the usual paper-and-pencil approaches and inkblot scores. Nor is it likely that peer-ratings of socially observable traits such as manifest anxiety, hostility, shyness, or dominance will have much in common with inkblot scores except in unusual circumstances. While such results are useful in pointing out certain kinds of inferences about the more superficial aspects of personality that it is unwise to make from inkblot scores, they are largely irrelevant to the broader issues of validity--developmental, cognitive and perceptual aspects of personality--as well as the psychodiagnostic evaluation of individuals with mental or emotional disturbances.

This conclusion is based on slim evidence, specifically:

Similar results were obtained by Ruebush (1960) and Barger and Sechrest (1961) in dealing with the more limited area of anxiety and hostility. Correlations between the two most relevant inkblot scores, Anxiety and Hostility, and such measures as Taylor's Manifest Anxiety Scale, Sarason's Test Anxiety Scale for Children, and Siegel's Manifest Hostility Scale--all rather similar paper-and-pencil approaches--proved to be insignificant. Barger and Sechrest also report a lack of any relationship between the inkblot scores and peer-ratings of anxiety and hostility, although correlations for the MMPI-type scores proved moderately significant.

The projective hypothesis notwithstanding, it is not unreasonable to suspect that these response set tendencies might also occur in projective measures. If Holtzman is correct, then social editing occurs only in objective measures. On the

other hand, Cronbach (1946, 1950) and all those who have followed feel editing occurs all inclusively. Is the Holtzman Inkblot Technique subject to the artifactual and confounding effects of response set? This is the question which this study is intended to investigate.

A simple contrasting of the HIT with known measures of response set would provide an incomplete answer to this question. It is possible that projective test information lies on a dimension far different from objective personality measures, but it does not seem unreasonable to suspect that response sets might exist in projective techniques in a unique and distinctive fashion, i.e., response sets unique to the HIT. This study has therefore a second goal, the goal of developing measures of response sets for the HIT that are clearly projective in nature and distinct from traditional objective personality-based measures of response sets.

## METHOD

In order to evaluate the effects of response sets on projective tests and to supplement the meager information pertaining to the scales, the following design was set up. A set of personality scales was selected representing major personality test response sets (SD, acquiescence, and the tendency to falsify) and attitude measures of SD and acquiescence.

## Scale Selection

Table 1 lists the scales chosen. Edward's SD (EDSD) and the MMPI second factor scale (R) represent the first two factors of the MMPI, respectively. Two scales, the MMPI Lie scale (LIE) and the Marlowe-Crowne SD scale (MCSD) were selected to represent the third factor of the MMPI because of their known similarity and the unclear nature of the factor. These MMPI factors are those delineated by Edwards, Diers, and Walker's 1962 article. The item overlap among these scales is not large and it was not considered serious because the intercorrelations of these scales have been established and are not overly high. Information on item overlap is presented in Appendix F.

The twenty HIT scales fall into three large and three small factors (individual administration). Holtzman et al. (1961) summarize the factor structure thus:

Table 1. Variables arranged according to their factor structure<sup>a</sup>

Name	Abbr.	I	II	III	IV	V	VI
MMPI FACTOR							
Selected Personality Scales <sup>b</sup>							
1 Edward's SD	EDSD	97	-18	01			
2 MMPI Second Factor	REP	07	-90	21			
3 MMPI Lie	LIE	25	61				
4 Marlowe-Crowne SD	MCSD	28	-55				
Attitude Scales							
5 Attitude SD (Taylor)	TASD						
6 Acquiescence (Peabody)	PA						
HIT FACTOR							
Holtzman Inkblot Scales, by factor <sup>c</sup>							
7 Integration	I	71	-11	12	-28	05	16
8 Movement	M	62	-08	52	00	-10	02
9 Human	H	68	-25	37	22	-02	-03
10 Popular	P	46	02	06	20	-15	-15
11 Form Definitness	FD	65	-30	25	35	-18	11
12 Barrier	Br	50	40	19	11	-22	14
13 Color	C	-13	60	22	-19	13	-20
14 Shading	Sh	-20	71	20	-10	-31	17
Form Definitness (reversed)	Fd-	65	-30	25	35	-18	11
15 Pathognomic Verbalization	V	06	05	52	-15	-24	-12
16 Anxiety	Ax	02	14	73	-10	05	-10
17 Hostility	Hs	07	04	89	-07	-12	14
Movement	M	62	-08	52	00	-10	02
18 Location	L	-31	-06	-21	77	01	-04
19 Form Appropriateness	FA	-18	19	-20	46	12	05
20 Rejection	R	-41	-21	-16	-07	09	-05
21 Animal (reversed)	A-	19	-09	06	08	-66	00
22 Space	S	-04	11	22	00	-19	14
23 Anatomy	At	-07	-05	-05	-19	25	35
24 Sex	Sx	06	10	01	-11	14	-16
25 Abstract	Ab	-02	21	40	-01	27	-33
26 Penetration	Pn	01	42	25	-20	03	41
New Scales							
27 HIT SD	HIT-SD						
28 HIT Acquiescence	HIT-A						
29 HIT Deviancy (general)	HIT-D						

<sup>a</sup>Decimals omitted.<sup>b</sup>From Edwards and Walsh, 1964.<sup>c</sup>From Holtzman et al., 1961.

Factor I. Invariably defined by Movement, Integration, Human, Barrier, and Popular, this factor usually accounts for more variance than any other. A high amount of this factor would be indicative of well organized, ideational activity, good imaginative capacity, well differentiated ego boundaries, and awareness of conventional concepts.

Factor II. Defined primarily by Color and Shading, and to a lesser extent by Form Definiteness, this bipolar factor involves sensitivity or responsiveness to the stimulus qualities of the inkblots. The positive pole of this factor would indicate over-reactivity to the color, shading, or symmetrical balance of the inkblot, while the negative pole would indicate primary concern for form alone as a determinant.

Factor III. Pathognomic Verbalization is the best single variable for defining this factor, although the loadings on Anxiety and Hostility are often higher. A high amount of this factor would be indicative of disordered thought processes coupled with an active, though disturbed, fantasy life.

Factor IV. Although not well defined because of the over-shadowing influence of Factors I, II, and III, in most samples, Form Appropriateness and Location serve as defining variables. The factor is bipolar, the positive pole tending to indicate perceptual differentiation coupled with a critical sense of good form, and the negative pole being indicative of immaturity, diffuse bodily preoccupations, and psychopathology.

Factor V. Reaction Time, Rejection, and Animal (reversed) are the primary variables defining this bipolar factor. Since there are no consistent trends involving variables other than these three, the factor is of minor importance and may emerge largely because of scoring dependence between Rejection and Reaction Time or Animal when the number of rejections is high.

Factor VI. Treated as a residual factor with respect to the patterns noted in the other five, this minor factor is defined largely by those scores which do not appear as marked variables anywhere else. Although the particular pattern of defining variables shifts considerably from one sample to the next, three variables stand out repeatedly --Penetration, Anatomy, and Sex.



Generally speaking, a high amount of this factor would be indicative of bodily preoccupations independent of similar components noted in Factors III and IV.

To date there has been one factor analytic study of the HIT that has included other measures of personality. Moseley Duffey, and Sherman (1963) report essentially the same HIT factor pattern outlined above. Their Factor II was composed entirely of objective personality measures and included both the MMPI SD scale and A scale (loadings of .87 and .76). The sample used was a clinical one; however, the authors took no cognizance of the possibility that their second factor could be viewed as a response set factor and not a content-meaningful one.

#### New Scales

Three new scales were created. The first was intended to parallel the personality scales SD factor and was called Holtzman Inkblot Technique -- Social Desirability (HIT-SD). There can be no such response as True (or False) to a inkblot, but it is certainly possible for groups of judges to rate a response for its social desirability-undesirability. This scale construction was undertaken with the view that, if successful, the measure would converge on other measures of SD or diverge if SD in the HIT were unique unto itself. HIT-SD was based upon the judgments of a group of raters (n = 20) judging the HIT responses as originally given. Thus

a supermatrix was created with the dimensions of 45 (the number of blots) X 29 (the sub-sample n) X 20 (the number of raters). However, the raters are not the same raters across the other two dimensions. Pragmatically this was a necessity. Eight hours would have been the minimal time requirement for each rater, something not obtainable when the raters are volunteers from general psychology courses. But the literature has adequately shown both the stability of rating procedures, and that 20 raters is a sufficient number for reasonable reliability. The score itself was obtained by first computing the mean rating given by the 20 raters for S one, blot one. The supermatrix was thus reduced to Ss X blots (29 X 45), with mean SD ratings in each cell. Rather than reduce the data further, the mean ratings were taken literally as evidence of SD responses; i.e., simply summed across the 45 blots. Each S in the sub-sample thus received a SD score based on the judgment of 20 raters, summed across 45 responses. Theoretically the range of such scores could be 45 to 405.

The second new scale follows the second personality scale factor in that it is a measure of acquiescence (HIT-A). Acquiescence in the HIT was defined after Dier's (1964) definition of acquiescence. A simple count was taken of the number of times a sub-sample S was given a SD rating in the range from 4.5 to 5.5. The theoretical range of such scores is 0 to 45.

The nature of the HIT-SD and HIT-A scoring procedures is such that they will correlate, but not necessarily highly.

Truncation of either score is a possibility, and indeed there is reason to believe that whatever truncation occurs, it will be toward neutral responses. Holtzman et al. (1961) indicate that, with group administration and college Ss, creative but not clinical responses are to be expected. A great number of everyday objects are seen, objects which would appear to be for the most part neutral; i.e., dogs, cats, trees, butterflies, landscapes, flowers, etc. The necessary interrelationship between these two scales should not pose a serious problem.

A third scale designed to detect general deviancy of responses from societal norms was constructed. Deviance was defined as any score on the 20 HIT variables that exceeded the 89th percentile of the scores obtained by a sample of University of Texas college students. (This sample is the most similar to the present sample among the many Holtzman reports). A "1" was assigned in such cases and a sum taken. The possible range was 0 to 20.

## PROCEDURE

## Administration and Subjects

The data were collected in two distinct phases. Phase one consisted of the administration of the HIT (group form) and the paper and pencil personality scales. Rating of the HIT responses for SD constituted the second phase.

All Ss were recruited from undergraduate courses in psychology and sociology. No age or sex limit was imposed, but graduates and non-citizens were excluded. The S pool for phase one was 171 and in phase two raters were run until the requirement that no single rater would rate the same response twice and a total of 20 raters for each blot was obtained.

Administration of phase one was by groups in a suitable auditorium. The auditorium was of the rising tiers of seats type to insure maximum visibility for all. Phase two was administered so that all raters could read the HIT response for a particular S to a particular inkblot, mark their choice and proceed to the next S's response to the same blot. The blot in question was projected onto a screen to provide a reference.

The Ss in both phases were instructed according to standard procedures (see Appendices A and B). It was necessary to add to the standard instructions for obtaining SD ratings in view of the fact that occasionally a HIT response

would be a reject, a non-response. Under these circumstances the raters were asked to rate, "What it means to not respond to this blot, to not see anything; though this (rating) may seem a bit difficult." All instructions were orally paraphrased and questions answered, except those which inquired about what the E thought was a good response or an appropriate rating.

In both phases, Ss were assured of anonymity. Raters were not allowed to observe the names of the Ss who had given HIT responses. All testing was done by E with administrative aid where necessary.

### Scoring

The six personality scales were scored according to the system of the original authors. Taylor's attitude measure of SD and Peabody's acquiescence measure were treated somewhat divergently from the authors' original systems.

Taylor (1961) originally had his Ss agree or disagree. This was expanded to a five point response scale (1, strongly agree; 3, neutral; 5, strongly disagree). Though the expansion of a scale generally has little effect, it was done in this case to avoid presenting the same item in two different forms. That is, there are five overlapping items between TASD and Peabody's acquiescence measure (PA). The PA requires the five point system, therefore the decision was made in its favor.

The PA scale is quite unusual in its scoring procedure. As stated previously, an acquiescent response is obtained only when a S double-agrees. Persons who agree to an item on the forward F scale and agree to the reverse of the item on the reversed F scale are scored as acquiescing to that item. All other combinations are scored zero. A simple sum is then taken to obtain the PA score. This scoring procedure diverges from the original only in that a six point scale was used. Again the two systems are compatible.

The HIT was scored for twenty variables (see Appendix C and Table 1) according to standard instructions. As in all cases, E performed the scoring; however, it should be pointed out that the E is not a professionally trained HIT scorer. Holtzman et al. (1963) feels that for research purposes this is acceptable though perhaps not with clinical populations or diagnostic purposes. It was assumed that whatever biases existent in E they were preferable to expanding the design to include other raters.

In order to assess the possibility that some correlations might be a spurious result of the E's scoring procedures a minimal check was performed. An independent scorer was trained by the E in the HIT scoring procedures. This second scorer then scored a selected random sample from the 580 individual HIT scores. Variables R and P were deleted because of their simple all or none scoring procedure. Variables S, Sx, and AB were deleted because of extreme truncation. The three

variables V, Ax and Hs were deleted because proper scoring does take a degree of "clinical intuition" and much practice. The remaining twelve variables were placed in a matrix with the sub-sample subjects, from which a sample of 20 was drawn at random. The interrater correlation between the author and an assistant over 20 randomly chosen scales was .98.

The full sample of 171 was too large in the pragmatic sense of finding sufficient raters to rate each response (a total of 7,695). Therefore, from this set of Ss a random, stratified sub-sample was taken. The stratification was based upon the Ss' obtained EDSO scores. The distribution of scores was split into six half standard-deviations from the mean ( $\bar{X} = 28.708$ ,  $\sigma = 6.1651$ ). This resulted in a selected distribution noted in Appendix E. Since the lowest five scores fall outside the lowest standard deviation, but constitute three per cent of the sample they were included with the lowest half standard-deviation. Thus this strata was seven per cent of the sample instead of four per cent. Each of the six strata were then sampled randomly according to its associated proportion (again see Appendix E for final sub-sample distributions).

### Hypothesis

The 29 scales detailed above can be regarded as falling into nine distinct sets of variables as shown in Table 2. Several potential patterns of correlations among these variables

are of importance. If HIT variables are unrelated to paper and pencil measures, variables in sets six through nine should have low correlations with variables in sets one through five. Low correlations between these sets could also occur because of specific non-contamination of the HIT variables with response sets. Separation of the above sources of low correlation is possible by examining set six, with its' four parts, against sets one through five and also against seven, eight and nine. If response sets do exist in projective methods but the method remains unrelated to paper and pencil measures, set six would be lowly correlated with sets one through five but sets seven, eight and nine would correlate highly with one through five. If set six showed low correlations with one through five and with seven, eight and nine the projective technique could be said to be uncontaminated with response sets of the standard personality variety or of the newly refined projective type. A strong relationship between set six and seven, six and eight, or six and nine would indicate a unique form of response set particular to the HIT. Sets seven, eight, and nine are of interest apart from any relationship with known measures of response sets. Correlations among the three should be low if they are indeed independent measures.

The purpose of this study is largely exploratory. Nevertheless, two facts seem compelling. Virtually no form of personality measurement in which extensive investigations have been conducted has been shown to be free of response set bias.



But general independence of different methods of assessment is also the rule. It is therefore hypothesized that the HIT is relatively free of response set influences as defined in MMPI type measures but that it is contaminated to some extent by response biases peculiar to itself and operationalized by one or more of the three new scales discussed above.

Table 2. Variables arranged according to method

Scale Number	Set
1	1. Personality SD
2	2. Personality Acquiescence
3,4	3. Personality Falsification
5	4. Attitudinal SD
6	5. Attitudinal Acquiescence
7,8,9,10,11,12	6. HIT (a) Normal
13,14,(-11),18,19	HIT (b) Self & environment perception
15,16,17,(18)	HIT (c) Pathology
20,21,22,23,24,25,26	HIT (d) Scoring and residuals
27	7. HIT SD
28	8. HIT Acquiescence
29	9. HIT Deviant

## RESULTS

The intercorrelations among variables were computed and are presented in Table 3. For the purpose of examination the table has been subdivided into six parts. The small upper left-hand triangle contains the correlations among the six personality and attitude measures. The large horizontal rectangle represents the intercorrelation between the personality and attitude variables and the HIT. The small rectangle in the upper right hand corner is the correlation between the six personality and attitude measures and the three new HIT scales. The large vertical rectangle contains the correlations between the new HIT scales and the regular HIT scales. The small triangle in the low-right corner is the intercorrelations among the three new scales. The large triangle represents the intercorrelations among the 20 HIT variables.

A significance level of .47 was chosen on the basis of 27 df and a P of .01. In Table 3 all correlations of .47 and higher are underscored. The value is from the tabled values of r in Steel and Torrie, 1960.

Sub-sample adequacy was partially answered by comparing the sub-sample to the sample on the 6 personality and attitude scales. Six separate t tests with n's of 29 and 171 showed no significant differences at the .01 level. This is not meant to be an assertion of the null hypothesis; rather, that

Table 3. Intercorrelations among the 29 variables<sup>a</sup>

	ED SD	LIE	Rep	MC Sd	PA	TA SD	R	L	S	FD	FA	C	Sh	M
EDSD	-	26	23	34	-16	-03	-09	20	-04	08	19	12	35	28
LIE		-	09	39	-11	-21	01	-15	-02	-31	-18	08	30	-05
Rep			-	-03	-38	14	-13	16	25	-14	14	00	-16	09
MCSD				-	-20	-06	20	03	-05	-20	-53	-15	-21	-21
PA					-	-44	39	-24	-33	19	06	-34	17	-03
TASD						-	-61	39	26	09	08	05	06	00
Rejection							-	-30	-14	-19	-12	-09	-21	-35
Location								-	-10	02	32	-22	06	06
Space									-	15	-14	00	-15	-21
Form Definiteness										-	26	-40	-49	49
Form Appropriateness											-	27	-01	32
Color												-	08	06
Shading													-	-32
Movement														-
Pathognomic Verbalization														
Integration														
Human														
Animal														
Anatomy														
Sex														
Abstract														
Anxiety														
Hostility														
Barrier														
Penetration														
Popular														
HIT-DEV														
HIT-Acq														
HIT-SD														

<sup>a</sup>Decimals have been omitted.

[illegible]

without evidence to the contrary the sub-sample probably does adequately reflect the full sample.

From Table 3 it is evident that the hypothesis of non-contamination of the HIT by personality or attitude response sets is demonstrated. All correlations in the HIT with personality and attitude scales are uniformly low. The three correlations above .47 might be considered chance occurrences. It is also apparent from Table 3 that the three new HIT scales bear little relation to the personality and attitude measures. Similarly the new scales bear little relationship among themselves. The relationship between HIT-SD and HIT-Acq was .18, a value lower than anticipated because of scoring procedures. The new HIT scales by standard HIT scales rectangle shows one strong correlation trend and two weak ones. HIT-SD and HIT-Acq are not systematically related to any HIT scales. The three correlations above .47 might again be attributed to chance. HIT-Dev has correlations above +.47 with scales FD, H, Ax, Hs and P. Additionally, scales M, I, Pn are +.40 or above and scale R is -.38. Thus the hypothesis of partial contaminations of the HIT with response set bias peculiar to itself is confirmed.

## DISCUSSION

The construct validity of the HIT is enhanced by the general non-relation between it and measures of response set found in this study. Bias free diagnostic tools are a clear and desirable goal and the HIT appears to approximate this goal. Specific projective bias does not exist insofar as HIT-SD and HIT-Acq were able to measure it. However, HIT-Dev does exhibit a distinct relationship to the HIT.

In examining the correlations between HIT-Dev and standard HIT scales, the expected negative correlation with Rejection ( $-.38$ ) is evident. This correlation is expected on the basis that the remaining 22 HIT scales correlate negatively with Rejection and the reported factor structures show a negative loading on all but one factor. Other scales with which HIT-Dev correlates also bear this same noticeable correlation: Human, Integration, Popular and Movement. The other correlations can be divided into two categories; those appearing on the HIT first factor (normal): Form Definiteness, Integration, Movement, Human, and Popular; and third factor clinical scales Anxiety, Hostility, and Penetration. Notably, V (Pathognomic Verbalization) is missing and Penetration is a sixth factor variable. Persons who score high on HIT-Dev also score high on the "normal" scales and the "clinical" scales but not on the most definite of the "clinical" scales (V). In short, such Ss give popular, precise, well

integrated responses involving humans in anxious, hostile situations and without specific pathological content. This does not seem a reasonable parallel to the tendency to falsify and/or to give pathological responses. It is perhaps more typical of creative persons who produce deviate responses but in normal, reality oriented fashion.

Conclusions about new measures have little meaning unless there is specific experimental evidence that the older measures were properly executed. The set of scales seven through 26 from Table 3 are the intercorrelations among the twenty HIT scales which in this case represent the evidence for adequate administration and scoring of the HIT. The correlations range from  $-.61$  to  $.80$  with twenty above  $.47$ . Though it is not true that visible clusters of correlations will accurately portray factor structure, in this case it is evident that a number of the factors found among HIT scores are present. Table 4 lists the intercorrelations among the six HIT scales comprising factor one. As is evident when correlations of this size are surrounded by correlations considerably smaller, the cluster is clear.

HIT factor two represented by Color, Shading and reversed Form Definiteness is unclear. Form Definiteness does correlate with Color and Shading ( $-.40$  and  $-.498$ , respectively) in the proper direction; however, the correlation between Color and Shading is  $.08$ , far too low.

The third factor, pathology, is reasonably clustered

except for Pathognomic Verbalization. In reviewing the item scores that contributed to V it is clear that most were of the Queer Response variety. A Queer Response is one in which, in the scorer's judgment, is sufficiently unusual in a pathological sense, but which lacks specific pathological content. Such responses rarely have sufficient anxiety content whereas other pathological responses often do. Thus at least a partial explanation can be offered for the low correlations with V. Variables in Factors IV, V and VI show low correlations. This leaves seven correlations above .47 unaccounted for.

Seven correlations above .47 from the remaining correlations is not unexpected on the basis of chance alone. However, a number of them are readily explainable in view of the nature of the sample of Ss. Rejection correlates negatively with all variables and most strongly with I (-.50), H (-.53), and P (-.61). Those Ss who responded with well integrated human popular responses tend not to reject blots. This is not surprising considering that many of the popular responses are about humans, and that Holtzman reports a loading of .41 for Rejection on the first factor, which contains scales Integration, Human, and Popular. As tempting as the explanation is, without further evidence it is very tentative in nature. Similar explanations for the remaining HIT intercorrelations do not appear possible. On the basis of the obtained factor structure, it is concluded that the scores



for the HIT are reliable and the same as those expected according to the literature.

The six correlations above .47 in the large rectangles might, as mentioned, be attributed to chance. However, several do have plausible explanations. The correlation of  $-.525$  between HIT-SD and R is in line with the strong tendency of R to be negatively correlated with everything. The  $-.61$  between T ASD and R can be viewed in a similar light. The  $-.48$  found between Sex and HIT-SD is reasonable in that the phenomenological impression that many of the raters gave E was that sex is not a very desirable thing to see in an inkblot or at least to write down. Most inquiries about what E thought was a SD or SUD response concerned sex. Cognitive ad hoc explanations for the remaining correlations exist, but are even more tenuous.

Table 4. Intercorrelations among HIT first factor and third factor variables

Factor I							:	Factor III			
I	M	H	P	FD	Br	:	V	Ax	Hs	M	
I	1.0	.47	.585	.66	.52	.41	V	1.0	-.03	.30	.24
M		1.0	.735	.61	.49	.28	Ax		1.0	.596	.42
H			1.0	.72	.80	.51	Hs			1.0	.55
P				1.0	.54	.27	M				1.0
FD					1.0	.40					
Br						1.0					

## SUMMARY

In this study an attempt was made to answer whether the Holtzman Inkblot Technique (HIT) is subject to artifactual and confounding effects of response sets. The study examined the correlation among three newly constructed measures of HIT specific response sets, the HIT and standard measures of response sets. The standard measures were: (a) Edwards' social desirability scale (EDSD); (b) MMPI Lie scale; (c) MMPI second factor scale (R); (d) Marlowe-Crowne social desirability scale (MCSD); (e) Taylor's attitude social desirability scale (TASD); and (f) Peabody's attitude measure of acquiescence (PA). HIT social desirability (HIT-SD), the first new scale, was defined for an individual S as the sum of mean social desirability ratings (from 20 judges) of that S's responses to 45 inkblots. The second new scale, HIT Acquiescence (HIT-Acq), was defined as the number of SD ratings that fell into the neutral zone. A 9-point absolute scale was used. The third scale, HIT-Deviancy (HIT-Dev) was defined as the number of HIT scores that exceeded the 89th percentile in a norms table for the HIT. It was hypothesized that the HIT would be free of response set influences as measured by MMPI type scales and would not be free of response set influences as measured by the new scales. The HIT was administered in the standard group form. Ss were drawn from introductory psychology courses. Of these a stratified random sample of 29 was taken.

The strata were based upon the EDSD scores. The 20 HIT scales were scored by E, with a check for scoring adequacy. The results showed the HIT to be unrelated to the response set measures EDSD, Lie, R, MCSD, TASD, and PA. HIT was found to be unrelated to HIT-SD and HIT-Acq. HIT-Dev did show a positive relation to first factor HIT scales and to HIT third factor scales. Five of these correlations were above the selected significance level of .47 and three additional correlations were above .40. Persons scoring high on HIT-Dev were characterized as being creatively deviate but not showing accompanying pathology. The scoring check was positive; inter-scorer correlation was .98. It is concluded that the hypothesis has been confirmed.

# REFERENCES

- Beck, S. J. The science of personality: nomothetic or idio-graphic? Psychological Review, 1953, 60, 353-359.
- Boring, E. G. A history of experimental psychology. New York: Appleton-Century-Crofts, 1950.
- Couch, A., & Keniston, K. Yeasayers and naysayers: agreeing response set as a personality variable. Journal of Abnormal and Social Psychology, 1960, 60, 154-174.
- Cowen, E. L., & Frankel, G. The social desirability of trait-descriptive terms: applications to a French sample. Journal of Social Psychology, 1964, 63, 233-239.
- Cowen, E. L., Staiman, M. G., & Woltizky, D. L. The social desirability of trait descriptive terms: application to a schizophrenic sample. Journal of Social Psychology, 1961, 54, 37-45.
- Cronbach, L. J. Response sets and test validity. Educational and Psychological Measurement, 1946, 6, 475-494.
- Cronbach, L. J. Further evidence on response sets and test design. Educational and Psychological Measurement, 1950, 10, 3-31.
- Crowne, D. P., & Marlowe, D. The approval motive. New York: Wiley, 1964.
- Diers, Carol J. Social desirability and acquiescence in response to personality items. Journal of Consulting Psychology, 1964, 28, 71-77.
- Edwards, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. Journal of Applied Psychology, 1953, 37, 90-93.
- Edwards, A. L. Social desirability and performance on the MMPI. Psychometrika, 1964, 29, 295-308.
- Edwards, A. L., Diers, Carol J., & Walker, J. N. Response sets and factor loadings on 61 personality scales. Journal of Applied Psychology, 1962, 46, 220-225.
- Edwards, A. L., & Walsh, J. A. Relations between various psychometric properties of personality items. Educational and Psychological Measurement, 1963, 23, 227-238.

- Edwards, A. L. & Walsh, J. A. Response sets in standard and experimental personality scales. American Educational Research Journal, 1964, 1, 52-61.
- Holtzman, W. H. Holtzman Inkblot Technique: administration and scoring guide. New York: The Psychological Corporation, 1958.
- Holtzman, W. H. Objective scoring of projective tests. In B. M. Bass and I. A. Berg (Eds.), Objective approaches to personality assessment. Princeton: Van Nostrand Co., 1959, Pp. 119-145.
- Holtzman, W. H., Moseley, E. C., Reiner, R. C. & Abbot, Elaine. Comparison of the group method and the standard individual version of the Holtzman Inkblot Technique. Journal of Clinical Psychology, 1963, 19, 441-449.
- Holtzman, W. H., Thorpe, J. S., Swartz, J. D., & Herron, E. W. Inkblot perception and personality. Holtzman Inkblot Technique. Austin Texas: University Texas Press, 1961.
- Iwawaki, S., & Cowen, E. L. The social desirability of trait-descriptive terms: applications to a Japanese sample. Journal of Social Psychology, 1964, 63, 199-205.
- Klett, C. J. The social desirability stereotype in a hospital population. Journal of Consulting Psychology, 1957, 21, 419-421.
- Klett, C. J., & Yaukey, D. W. A cross-cultural comparison of judgments of social desirability. Journal of Social Psychology, 1959, 49, 19-26.
- Lentz, T. F. Acquiescence as a factor in the measurement of personality. Psychological Bulletin, 1938, 35, 659.
- Lorge, I. Gen-like: halo or reality? Psychological Bulletin, 1937, 34, 545-546.
- Lövaas, O. I. Social desirability ratings of personality variables by Norwegian and American college students. Journal of Abnormal and Social Psychology, 1958, 57, 124-125.
- Mathews, C. O. The effect of the order of printed response on interest questionnaire. Journal of Educational Psychology, 1929, 34, 128-134.
- Moseley, E. C., Duffey, R. F., & Sherman, L. J. An extension of the construct validity of the Holtzman Inkblot Technique. Journal of Clinical Psychology, 1963, 19, 186-192.

- Peabody, D. Attitude content and agreement set in scales of authoritarianism, dogmatism, anti-semitism, and economic conservatism. Journal of Abnormal and Social Psychology, 1961, 63-1, 1-11.
- Peabody, D. Authoritarianism scales and response bias. Psychological Bulletin, 1966, 65-1, 11-23.
- Rorer, L. G., & Goldberg, L. R. Acquiescence in the MMPI? Educational and Psychological Measurement, 1965, 25-3, 801-817.
- Steel, R. G., & Torrie, J. H. Principles and procedures of statistics. New York: McGraw-Hill, 1960.
- Swartz, J. D., & Holtzman, W. H. Group method of administration for the Holtzman Inkblot Technique. Journal of Clinical Psychology, 1963, 19, 433-441.
- Taylor, J. B. Social desirability and MMPI performance: the individual case. Journal of Consulting Psychology, 1959, 23, 514-517.
- Taylor, J. B. What do attitude scales measure: the problem of social desirability. Journal of Abnormal and Social Psychology, 1961, 62-2, 386-390.
- Watson, R. I. Historical review of objective personality testing: the search for objectivity. In B. M. Bass & I. A. Berg (Eds.) Objective approaches to personality assessment. Princeton, N. J.: Van Nostrand, 1959. Pp. 1-23.
- Wood, B. D. Studies of achievement tests. Journal of Educational Psychology, 1926, 17, 1-22.
- Zubin, J. Failures of the Rorschach technique. Journal of Projective Techniques, 1954, 18, 303-315.

## ACKNOWLEDGMENTS

I would like to express my gratitude and appreciation to Dr. James A. Walsh, co-chairman of my dissertation committee and major professor, for his patient effort, support and guidance in making this dissertation a reality.

I would also like to thank Dr. Wilbur L. Layton, committee co-chairman, for his patience, interest and support; and Dr. David H. Mills, Dr. Thomas F. Lyons, and Dr. Neal A. Palomba for their guidance, interest and support.

## APPENDIX A

## Instruction Sheet

Directions: Below are four responses that a person might give when asked to say what he sees in an inkblot.

1. A horse's head
2. A man stabbing another man.
3. Humpty Dumpty carrying something.
4. A mother and child walking.
5. A mother helping her child with his shoe.

What we would like you to tell us is how socially desirable or undesirable it is to GIVE a particular response. We are not interested in whether you would respond in this way; but rather, how socially desirable or undesirable you think it is for people in general to GIVE a particular response to an inkblot. Please use the scale below to show your choice.

<u>Choices</u>	<u>Meaning of Choices</u>
1	Extremely Undesirable
2	Strongly Undesirable
3	Moderately Undesirable
4	Mildly Undesirable
5	NEUTRAL
6	Mildly Desirable
7	Moderately Desirable
8	Strongly Desirable
9	Extremely Desirable

Remember that you are to judge the responses in terms of whether you consider them desirable or undesirable in others.

PLEASE DO NOT SKIP ANY RESPONSES.

(Verbally Ss were instructed on rejected or blank responses and to disregard second responses if and when they occurred. Also, they were told that the task was not to try and see what other people said they saw, nor should they give strictly ideosyncratic responses. That is, if they dislike apples they should not judge on this basis but how the generalized other would rate.)



## APPENDIX B

## Instruction Sheet:

"You will be shown a series of inkblots, each of which will be projected on the screen before you for one minute or so. Using your imagination, write down in the space provided a description of the first thing the blot looks like or reminds you of."

"Include in your description the particular characteristics or qualities of the inkblot which are important in determining your responses--i.e., what about the blot made it look that way? Give as complete an answer as you can in the time available."

"None of these inkblots has been deliberately drawn to look like anything in particular. No two people see exactly the same things in a series of inkblots like these. There are no right or wrong answers."

## Things to Remember:

"Outline or trace the area of the blot used."

"What do you see in the inkblot?"

"What about the blot makes it look that way?"

## APPENDIX C

Variables from the Holtzman Inkblot Technique (HIT)<sup>1</sup>

Form Definiteness - The definiteness of the form of the concept reported, regardless of the goodness of fit to the inkblot. The greater the score, the more definite the concepts are in form.

Form Appropriateness - A 3-point scale measuring the goodness of the form of the percept to the form of the inkblot. The higher the score, the better the fit.

Reaction Time - The time, in seconds, from the presentation of the inkblot to the beginning of the primary response.

Rejection - The number of inkblots for which the individual does not give a scorable response.

Space - A variable indicating the degree to which the white part of the card is used as the figure and the inkblot as background.

Balance - A variable indicating a concern with the symmetry or asymmetry of the inkblots.

Color - A 4-point scale measuring the apparent primacy of color (including black, gray, or white) as a response-determinant. The higher the score, the more color is used as a primary determinant of the response.

Shading - A 3-point scale measuring the apparent primacy of shading as a response determinant. The higher the score, the more shading is used as a primary determinant of the response.

Movement - A 5-point scale measuring the amount of movement or potential movement the subject voluntarily ascribes to the percept. The higher the score the more the movement.

Integration - A 2-point scale indicating that the subject has organized two or more adequately perceived blot elements into a larger whole. The higher the score the greater the integration.

Human - A 3-point scale indicating human content seen in the inkblots. Each blot is scored zero for no human content, 1 for parts of a human, and 2 for whole humans.

---

<sup>1</sup>This table is reproduced from Moseley, et al., 1963.

Animal - A 3-point scale indicating animal content seen in the inkblots. Each blot scored zero for no animal content, 1 for parts of animals, and 2 for a whole animal.

Anatomy - A 3-point scale indicating anatomical content seen in the inkblot. The greater the score, the more visceral and the cruder the anatomical content.

Sex - A 3-point content scale ranging from no sex references to blatant sex references.

Abstract - A 3-point scale for coding the degree of abstractness of the concept. The higher the score, the more abstract the concept.

Anxiety - A 3-point scale derived from scoring the content for signs of anxiety, such as expressed or implied emotions and attitudes; expressive behavior, and cultural stereotypes of fear. The higher the score the more the content indicates anxiety.

Hostility - A 4-point scale used to scale content for symbolic, implicit, or explicit or actions to more direct, violent ones, the score increases.

Barrier - A 2-point scale used to score content for reference to any protective converging membrane, shell, or skin that might be symbolically related to the perception of body image boundaries.

Penetration - A 2-point scale reflecting any concept which might be symbolic of an individual's feeling that his body exterior is of little protective value and can be easily penetrated.

Location - The size of the area used relative to the total area of the inkblot. In general, the greater the area used, the lower the score.

Verbalization - A variable in which nine qualitatively different kinds of autistic responses are merged into a single scale. Each inkblot is scored from zero to 4, and a high score reflects increased autism and pathology.

Popular - A specified common response to one of 25 inkblots in each form of the HIT. The greater the score, the more the subject tends to see popular percepts.

## APPENDIX D

The Standard Group Method of Administration<sup>1</sup>

Description: In its present standardized form, the group method involves several modifications in the instructions and procedures as employed for the individually administered version of the HIT. The administration of the group method consists of the following procedure:

1. The directions on the front of the special group booklet (or printed on individual sheets when the individual Record Form is used in the group method) are read aloud to the subjects while they are reading them silently.

"You will be shown a series of inkblots, each of which will be projected on the screen before you for one minute. Using your imagination, write down in the space provided a description of the first thing the blot looks like or reminds you of."

"Include in your description the particular characteristics or qualities of the inkblot which are important in determining your responses--i.e., what about the blot made it look that way? Give as complete an answer as you can in the time available."

"None of these inkblots has been deliberately drawn to look like anything in particular. No two people see exactly

---

<sup>1</sup>This section is reproduced from Holtzman, 1958.

the same things in a series of inkblots like these. There are no right or wrong answers."

2. Trial inkblot X is projected on the screen. The examiner states that a common response to this inkblot is a "bat or winged creature." He outlines on the screen the area of the inkblot used in this response (W, omitting the d's on each side), points out the various parts (head, wings, tail) of the "bat," and briefly mentions the role of form in determining this response. The examiner then explains that such a response might be written as "bat because of the form" in the space provided, outlining the appropriate area on the accompanying diagram. Next, the examiner says that another common response to this inkblot is "pool of oil," which serves to illustrate the use of color and shading rather than form as determinants. Finally, the examiner states that still another response to this inkblot is a "steer's head," outlines the area used (center D), and discusses the roles of form, color, and shading in determining it.

Card Y is then projected on the screen, and the examiner points out a common response, "human figure", and mentions the role of form in determining the response. Using the same area, he then cites another common response, "skeleton", and points out the role of form and shading as determinants. Finally, the examiner states that still another response to this inkblot is "blood", thus providing an opportunity to illustrate a response with color alone as the primary

determinant.

3. The initial instructions are repeated (paraphrased), and the subjects are asked if they have any questions.

4. In accordance with a prearranged schedule during the test series, the subjects are occasionally reminded of important aspects of the instructions. Eight inkblots scattered through the series are verbally reinforced as indicated below:

Card Number	Verbal Reinforcement
2	"Write out as complete a description as you can in the time and space available."
3	"Just let your imagination run, and put down what the inkblot suggests to you--what you see in it."
6	"This is another one of those blots where you'll have to be careful in outlining that part of the area which you use."
8	"Write out as best you can what characteristics of the inkblot were deciding factors in your response."
9	"Be sure to draw a line around that part of the blot that suggested your response."
14	"We're particularly interested in knowing what aspects of the inkblot influenced your response."
19	Same as for Card 9.
24	Same as for Card 2.

5. Inkblots 1, 2, and 3 are each exposed for 120 seconds; 4, 5, and 6 for 100 seconds; 7, 8, and 9 for 90 seconds; and the remaining thirty-six inkblots are each exposed for 75 seconds.

In the standardized group method of administration for the HIT, the scoring of two inkblot variables--Balance and Affect Arousal -- has been discontinued because of the

difficulties in scoring it accurately.<sup>1</sup>

---

<sup>1</sup>Because of the serious difficulties in scoring Affect Arousal, the variable also is no longer scored routinely in the individual version of the HIT.

APPENDIX E



# Sampling Statistics for Sub-Sample

$\bar{x}=28.708$ $s^2 = 6.1651$ $s^2/2 = 3$ $n = 29$									
Actual # Ss chosen	2	2	3	5	6	6	4	1	
# of subsample Ss' by %	1.23	2.45	3.33	4.93	6.10	6.30	3.7	1.05	
% of Total	(Combined)								
	02.9	04.1	08.2	11.1	16.4	20.5	21.1	12.3	03.5
Total # of Ss appearing in a half standard deviation	5	7	14	19	28	35	36	21	6
(below 16) (17-19) (20-22) (23-25) (26-28) (29-31) (32-34) (35-37) (38-40)									

## APPENDIX F

Items held in common

	SD	LIE	Rep	MCSB	TASD	PA
SD	-	0	1	0	0	0
LIE		-	0	2	0	0
Rep			-	0	0	0
MCSD				-	0	0
TASD					-	5
PA						-