Roraszchach O Responses and the Concept of Originality

Joseph Federico
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses

Recommended Citation
https://ecommons.luc.edu/luc_theses/2733

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 1974 Joseph Federico
ACKNOWLEDGEMENT

Many people have earned a special notice of appreciation for providing needed help at appropriate times. In particular, I would like to acknowledge the unfailing encouragement and the many instances of daily support provided by my wife, Ann. In addition, much substantive support at many levels was provided at a number of crucial points by Frank J. Kobler, Ph.D. Finally, the contributions of Robert Nicolay, Ph.D. and of Kim Poggenpohl, in particular, and of the volunteer examiners and voluntary test subjects were also essential to the successful completion of this project.
VITA

Joseph Federico was born on August 12, 1938, in New York, N.Y. He finished grade school in New York City, junior high school in Yonkers, N.Y., prep school in Hebron, Maine, and college in Easton, Pennsylvania. He received a M.S. degree in General Psychology from the University of Oklahoma and moved to Chicago in 1964 to attend the Ph.D. program in clinical psychology at Loyola University.

Mr. Federico is married, with three children, and lives in Evanston. He completed a psychology clerkship at the West Side Veterans' Administration Hospital in Chicago, and a two-year clinical internship at the Charles F. Read Zone Center in Chicago. Upon completion of the internship, he became Chief of Services of an inpatient unit at the Read Zone Center. Following that, Mr. Federico went to work for the Mental Health Division of the City of Chicago Board of Health at the Lakeview-Uptown Mental Health Center as a clinical psychologist. After nearly two years there he became Director of the Northtown-Rogers Park Mental Health Center, a new city-sponsored facility. He remained there for 18 months and then began working at the Edgewater-Uptown Mental Health Center. He was Assistant Director in charge of four service outposts; an around-the-clock, 365-day-per-year Emergency Service element; and the Inpatient element.

Mr. Federico has worked part-time at the Catholic Family Consultation Service as a marital counselor. He has also worked part-time at St. Joseph's Carondelet Child Care Center. He presently works part-
time at the Ravenswood Hospital Community Mental Health Center as a Crisis worker on the 24-hour-per-day Emergency Service.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgments</td>
<td>ii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>The Problem</td>
<td>3</td>
</tr>
<tr>
<td>Purpose</td>
<td>10</td>
</tr>
<tr>
<td>II. Review of Related Literature</td>
<td>13</td>
</tr>
<tr>
<td>Research on Rorschach O Responses</td>
<td>13</td>
</tr>
<tr>
<td>Research on the Embedded Figures Test</td>
<td>24</td>
</tr>
<tr>
<td>Research on the Remote Associates Test</td>
<td>29</td>
</tr>
<tr>
<td>III. Methodology</td>
<td>34</td>
</tr>
<tr>
<td>Subjects</td>
<td>34</td>
</tr>
<tr>
<td>Examiners</td>
<td>35</td>
</tr>
<tr>
<td>Tests and Procedures Used</td>
<td>36</td>
</tr>
<tr>
<td>Tabulations and Statistics</td>
<td>39</td>
</tr>
<tr>
<td>IV. Results and Discussion</td>
<td>48</td>
</tr>
<tr>
<td>Results</td>
<td>48</td>
</tr>
<tr>
<td>Discussion</td>
<td>51</td>
</tr>
<tr>
<td>V. Summary and Conclusions</td>
<td>68</td>
</tr>
<tr>
<td>Summary</td>
<td>68</td>
</tr>
<tr>
<td>Conclusions</td>
<td>72</td>
</tr>
<tr>
<td>References</td>
<td>77</td>
</tr>
<tr>
<td>Appendix A</td>
<td>82</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Scoring of Male Q. Thlta by Two Judges Using Hertz Tables</td>
<td>40</td>
</tr>
<tr>
<td>2.</td>
<td>Scoring of Female Q. Data by Two Judges Using Hertz Tables</td>
<td>41</td>
</tr>
<tr>
<td>3.</td>
<td>Total Test Scores of Male Subjects</td>
<td>42</td>
</tr>
<tr>
<td>4.</td>
<td>Total Test Scores of Female Subjects</td>
<td>43</td>
</tr>
<tr>
<td>5.</td>
<td>Rank-Order Intercorrelations Between Rorschach Measures and the Embedded Figures Test (EFT) and the Remote Associates Test (RAT) for Male Subjects</td>
<td>45</td>
</tr>
<tr>
<td>6.</td>
<td>Rank-Order Intercorrelations Between Rorschach Measures and the Embedded Figures Test (EFT) and the Remote Associates Test (RAT) for Female Subjects</td>
<td>46</td>
</tr>
<tr>
<td>7.</td>
<td>Rank-Order Intercorrelations Between Rorschach Measures and the Embedded Figures Test (EFT) and the Remote Associates Test (RAT) for All Subjects</td>
<td>47</td>
</tr>
</tbody>
</table>
CHAPTER I

STATEMENT OF THE PURPOSE

Introduction

This study will deal with the subject of Rorschach original (Q) responses. An Q response has been defined as a response that occurs one time or less per 100 responses to a particular blot area (Rorschach, 1942). The number of Qs per protocol are routinely recorded and summed as part of the Q scoring category. This scoring category provides a test examiner with a quantitative measure of Q production by each subject. These data are presumed to communicate useful information to the examiner, who can theoretically put the data to some practical clinical use. However, it is not inconceivable that Rorschach psychologists are in fact collecting data that have no real clinical relevance. One purpose of this study is to investigate the possible clinical significance of Rorschach Q responses.

A Rorschach clinician's first inclination is to evaluate the number of Q responses that are obtained on each individual protocol and to determine where along some normal-abnormal continuum the number falls; and then to integrate that determination into a number of other conclusions that had previously been made about the subject. The result of this would be a testing report in which Q data would take their place as an integral and contributing element.

The fact of the matter is that this happens rarely, if at all.
Q data are rarely incorporated into test write-ups, and if they are they do not fit readily into the body of the text. Statements about Q totals are usually set apart from statements about the data from other scoring categories because the meaning of Q data has never been explained in relation to data from the other categories (except occasionally, and then the discussion is in terms of "the P-Q dimension" - which sets both P and Q apart from the rest of the scoring interpretations). The reason for this probably is because there is no substantiated clinical inferences that can be validly based upon Q totals. No one seems quite sure what a given number of Q responses means.

An interest in investigating the clinical significance of Q responses is one source of motivation for this study. A second major interest underlying this study is methodological. Any research involving Q responses is bound to raise a number of interesting methodological issues. Just scoring Qs is an issue; determining what is an average range of total Qs per record is an issue; deciding whether or not to make or maintain a distinction between Q+ and Q- responses is an issue; assuring sufficient Q responses to satisfy the assumptions underlying the use of various statistical procedures is an issue, and so on. There are many challenging methodological considerations that seemingly must be confronted and resolved if the topic of Q responses will stimulate more research activity. Some of these issues alone seem centrally related to an understanding of the basic nature of Q responses.

A third major interest underlying this study is a wish to
investigate and hopefully to get an increased understanding of the concept of originality. In many ways originality is a fascinating concept. It is at once definable and necessarily loose. Originality can be defined in terms of a strictly quantitative statistical criterion or it can be defined in terms of more qualitative considerations such as usefulness, appropriateness to the surrounding context, etc. Originality can run a gamut from highly, or totally, idiosyncratic to task-oriented creativity. There is an obvious question regarding the standard by which originality is defined, and how to understand the obvious similarity between originality and creativity. Finally, there is a question about the nature of the psychological processes that seem to underlie original psychological products. At this point the discussion has come full circle, for this feeds directly into the first major interest of this study.

The Problem

The most basic problem for this study is that there are so few previous studies that have dealt with the subject of Q responses. This one problem precipitates a string of subsequent problems. The scarcity of literature on the subject means that there are virtually no conceptual or methodological precedents, no guidelines, and no possibilities which can be confidently discarded. This means, in turn, that any research must be conducted in the face of seemingly unlimited possibilities. This situation presents many problems for the present study. A major interest of this study is to investigate the nature of the psychological processes that are measured by the Q category. As matters stand, that is a totally unique problem. There are no actuarial
data regarding the incidence of Q responses; there are no studies that suggest the underlying dynamics of Q-giving; there is no compelling reason to accept the clinical maxim that Q responses represent a predisposition to original thinking; there is no generally accepted theoretical framework for understanding the concept of originality. Instead, there are a set of briefly-stated assumptions drawn from observation and carried on by tradition, and these have long served as guidelines for understanding Q responses.

The problem that this presents is that these observations may or may not be true - no one is quite sure and no one can quite argue the point. The lack of data hampers the very debate that the presence of pure clinical speculation fosters. Thus, the examiner who begins to wonder about the validity on the usefulness of the clinical significance of a given number of Q responses or any record is as unsure about what to do with his wonderment as he is about what to do with his data. This problem becomes one of defining a starting place from among the numerous possibilities that exist. The solution that was employed here in response to this problem was to use some combination of logic and empirical observation as a basis for formulating beginning inferences about how to proceed in investigating the issues of interest.

For instance, after testing it was often noted that high numbers of Qs were obtained from evidently well-adjusted test subjects while low or average numbers were obtained from clearly disturbed subjects. Also, there seemed to be a higher median number of Qs from all subjects as a group than one would expect from guidelines reported in projective
texts. Logically, it made sense that this should be so since each individual may be assumed to produce a moderate but definite number of idiosyncratic (i.e., original) responses. It seems to follow from this that the percentage of Q responses per record would be a more meaningful measure than the absolute number of Q responses. This is because the percentage takes into account the total number of responses per record while the absolute number of Qs does not.

Approaching the issue in this manner quickly leads to many innovative ideas. In the process it also challenges many of the traditional but untested assumptions (conclusions) about Q responses. One of these has to do with the very basic matter of the definition of the term "original". Where Q responses are concerned, originality is defined in strictly quantitative terms. Such a definition tends to ignore any consideration of the qualitative aspects of responses. It equates originality with uncommonness. In this sense the definition does have certain drawbacks, but on the whole it appears to provide the basis for a sound operational definition of a psychological concept. The definition would be considered acceptable on this basis if there was some reliable way to determine when a response met the one-in-one-hundred criterion for originality. In other words, a definition based on a solely quantitative criterion would be considered acceptable if there was some objectively verifiable means of checking whether or not any given response meets that standard.

To consider such a definition acceptable is to immediately say two additional things. One is that the traditional method of scoring Q responses is unacceptable. The traditional method is to have the
test examiner assign an \( Q \), no-\( Q \) scoring based entirely on his memory of the past responses given to any specific blot area during tests that he administered. The idea that every clinician would be able to remember which responses had been given more than, equal to, or less than 100 times per blot area is unrealistic on many counts. It is unreasonable to expect an examiner to remember the content of each response given by hundreds of subjects; and it is unrealistic to expect an examiner to remember the exact limits of the blot areas associated with the response content; and it is unrealistic to expect that all examiners will encounter equivalent testing populations.

The second additional consideration that is raised by accepting the idea of a quantitative basis for defining originality is that frequency tables of Rorschach responses are needed to score \( Q \)s. This would involve testing various identifiable subject groups (children, adolescents, adults, hospitalized psychotics, etc.) and presenting all the responses obtained, along with the appropriate scoring, in frequency tables. Such tables could be used as standards for making reliable scorings and in conducting replicable research concerning \( Q \) responses. It happens that there are a number of tables of this nature as part of the Rorschach literature. One set of tables was published by Hertz (1951) and presents the Rorschach responses of 850 adolescents, ages 11-16 and of high average intelligence, from the Cleveland school system. Another set of tables, published by Small (1956), presents more than 6,000 Rorschach responses scored for area, determinant, and content by Beck and 17 other Rorschach workers. There is also a two-volume set, published by Thomas, Ross, and Freed (1964,
1965), that presents the Rorschach responses of 586 medical students.

It will be noted that each of the published sets of tables is essentially unrelated to the others and addresses itself to some unique aspect of Rorschach testing. The tables published by Small and by Thomas, et al., do not include Q or P scorings, for instance, which makes it impossible to relate them to the Hertz data along those dimensions. This means that the kind of frequency tables that might be most desirable do not exist and that any attempt to select one of the existing ones as an objective scoring standard will involve making a choice between alternatives that have secondary desirability. With this understanding, the Hertz tables are considered the best of the readily available objective scoring standards for Q responses.

The Hertz tables are the only ones that contain Q scorings, and although all the responses are from adolescent subjects it is reasonable to assume that from a psychometric standpoint adolescent students of high average intelligence are probably a reasonable equivalent of an unselected adult population. Evidently this same logic was employed by Hertz and Paolino (1960) in a study where they compared the Rorschach responses of a group of psychotic adults. Their study used the Hertz tables as the scoring standard, and it involved scoring Q responses as well as developing an "original score" for each subject. The use of the Hertz tables in this study suggests that the responses of teen-age school children comprise acceptable criteria for scoring the Rorschach responses of adult subjects. This precedent, plus the fact that no other set of published frequency tables include Q scorings, are the reasons for accepting the Hertz tables as the
standard for scoring Q responses in the present study.

Once an objective standard for the Q scoring category is accepted, serious attention can be directed toward the question of the nature of the psychological processes involved in the development of Rorschach Q responses. There are two immediately apparent ways of beginning to investigate this question. One is to review numbers of Q responses that have already been given. The purpose of this would be to see if there are any common elements of these responses that seem to characterize the Q-giving process. The second way is to examine the Rorschach literature for theories regarding the genesis of Q responses. This study has utilized both these approaches in combination.

The first approach involved the examination of 40 randomly selected Rorschach protocols that the experimenter had in his own files, plus the responses scored Q in the Hertz tables. This procedure led to a conclusion that will serve as a basic assumption for the present study, namely that Rorschach Q responses seem to result from one of two discriminably different processes. The first process involves giving almost any response to an infrequently-seen area of the blot. In this type of response, the area of the blot that is used is so rare that virtually any response will earn an Q scoring. It is the use of that particular area of the blot that meets the one-in-one-hundred criterion, and the content of the response is almost incidental.

The second type of process produces a response in which the content is of primary importance. In this type of response the subject responds to a frequently-seen blot area with a percept that
is reported once in one hundred or more times. Here, it is the assignment of a highly idiosyncratic (i.e., infrequent) meaning that defines the response as original. The blot area is not original, while the response content is original.

This distinction seemed to be so clear-cut when examining the Q responses that there seemed to be little doubt about the existence of two inherently different psychological processes. In fact, separate names for each of these processes were created in order to acknowledge their seemingly separate realities. The first (Type One) process, involving the assignment of meanings to rarely seen blot areas, was termed idiomorphic. The second (Type Two) process, involving the assignment of uncommon meanings to frequently-seen blot areas, was termed idiographic. Thus, empirical examination of actual Q responses scored according to the Hertz tables led to the working hypothesis that there are two discriminably different processes, termed idiomorphic and idiographic, that underlie the production of Rorschach Q responses.

The second approach was to examine the psychological literature for discussions of either Rorschach Q responses or the general concept of originality. This approach revealed that there was very little literature on the topic of Rorschach Q responses, and equally little literature dealing with the concept of originality. Actually, the literature on originality tends to overlap with the literature on creativity, but the literature on creativity was considered inapplicable to this study because the Rorschach test does not have a scoring category for creative responses. The focus of this study is originality as
defined by Rorschach usage, and this must be clearly distinguished from the concept of creativity. This distinction shall be strictly maintained throughout this study, with the term originality not to be synonymous or interchangeable with the term creativity.

This study will take the position that Maltzman expressed in one of the few articles on originality in the psychological literature. Maltzman states:

Originality...refers to behavior which occurs relatively infrequently, is uncommon under given conditions, and is relevant to those conditions. Creativity...refers to products of such behavior and the reactions of other members of a society to those products. Our distinction implies that an individual may be highly original but not creative. A great many more behavioral and societal variables influence creativity than originality, making the study of originality under simplified laboratory conditions more feasible than that of creativity (1960, p. 229).

Aside from pointing up the issue of maintaining a clear distinction between the concepts of originality and creativity, the search of the literature did little to help define or sharpen the problem of this research. The literature provided little substantive information about originality and Rorschach Qs. Its major value at this point was in indicating that the question itself was original and that any research of this nature would break new ground. This is the reason that the basic problem was arrived at primarily through practical experience and empirical observation.

Purpose

The overall purpose of this study is to examine the Q category in its own right in order to understand its nature and possible significance as a psychological measure. In order to do this the study was conceptualized in terms of three issues: first, the clinical
significance of the $Q$ scoring category; second, exploring methodological issues and developing appropriate methodological procedures; and, third, investigating the nature of originality as a psychological phenomenon.

Although these purposes can be stated separately for ease of discussion, there are ways in which they inevitably overlap. The most obvious way is in assuming that to know that a test subject has given a certain number of Rorschach $Q$ responses is to be able to make valid inferences about the subject's customary psychological functioning. Actually, such an assumption is also a very common one - so common, in fact, that one purpose of this study will be to put it to the test. This means that a further purpose of this study will be to investigate questions involving combinations of the three issues that have been presented separately up to this point.

In order to achieve these purposes, certain kinds of data must be obtained. The Rorschach Inkblot Test, the Embedded Figures Test, and the Remote Associates Test will be used in an attempt to generate data to be used in determining whether $Q$ responses reflect either or both of two distinctive kinds of psychological processes. The data obtained by these instruments will be used to test the following hypotheses:

1. There will be a uniform relationship between scores on the Embedded Figures Test and the total number of Rorschach $Q$ responses given by each subject.

2. There will be a uniform relationship between scores of the Remote Associates Test and the total number of Rorschach $Q$ responses given by each subject.
3. There will be no uniform relationship between the scores on the Embedded Figures Test and the number of correct responses given to the Remote Associates Test for each subject.
CHAPTER II

REVIEW OF THE RELATED LITERATURE

Research on Rorschach Q Responses: The Rorschach literature that is most pertinent to this study comes from two distinctly different sources, the theoretical and clinically-oriented literature and the empirical and experimentally-oriented literature. The clinical literature first assumes and then asserts a number of untested beliefs regarding the occurrence and clinical significance of Q responses. More specifically, the tests on projective techniques and the testers' handbooks accept as truisms the following assumptions: Qs are infrequent occurrences; they result primarily from a statistically uncommon content response to a frequently seen blot area; they can be satisfactorily scored by means of a method that uses the individual tester's memory of his past testing experience as the major scoring standard; and a subject with a demonstrated capacity for originality or resourcefulness will produce uniformly high numbers of Q responses to Rorschach stimuli.

The first mention made of Q responses in the psychological literature occurred in Hermann Rorschach's historical monograph. In that monograph, Rorschach (1942) mentioned original responses explicitly and suggested originality as a separate scoring category. This practice that he suggested has been carried into the present, although the clinical significance of Q responses remains an open question.
In his monograph, Rorschach suggested that a response be defined as original if it occurs once or less in one hundred responses to a given blot area. This definition has been accepted by subsequent projective testers (cf., Holt, 1968; Klopfer, et al., 1954; Kobler, 1964), but it presents a number of obvious problems.

One problem is in knowing when a given response meets the 1:100 standard. The question seems to become one of validating the scoring against some objective criterion. In the absence of such a criterion each tester must use his own experience as a standard. This immediately suggests that testers who have most of their experience with specialized populations - like hospital inpatients, jail inmates, outpatients in private practice, mentally retarded children, etc., - may be drawing upon discriminably different bodies of experience. This may be invalidating someone's scoring - alternatively, of course, it may be invalidating no one's scoring, or everyone's. The point is that no one can be sure because Rorschach's suggestion has not been followed up by research. Indeed, the issue of an objective standard by which to validate Q scoring seems never to have been raised in any research context.

Even Anne Roe (1952a), who did a series of studies that present the most comprehensive body of Q data in the literature, scored the Qs herself and used her past experience as her scoring standard. She administered group Rorschachs to groups of physical scientists, biologists, anthropologists, and psychologists. The mean number of total test responses for each group was 32.9, 32.7, 44.5, and 31.7, respectively; and the mean Q percent for each group was 18.6, 26.3,
8.7, and 7.8. She stated,

Scoring of original responses is one of the more subjective parts of all Rorschach scoring, but all of these were scored by me (as well as by an assistant) so any bias I may have as a psychologist or as a person should be a relatively constant factor. On the other hand, the drop is in the sequence in which the scoring was done, and it is possible that this has affected our criteria. There is some association here with 'technical responses', which are practically non-existent among psychologists, rare among anthropologists, frequent among physicists and very common among biologists. (Some of these are originals, but if given by a number in the group they were not scored as originals.) (1952a, p. 220)

What Roe is reporting in this quotation is that she scored the records using the scoring method recommended by the projective texts and in so doing she encountered the subjectivity problem. She tried to control for this by using herself as a baseline scoring standard but that did not work so well because she was dealing with different subject groups, each of which had its own unique group characteristics. Roe then switched to using each separate group as its own baseline scoring standard, which is a better procedure although still troublesome. She herself notes that determining which responses are Qs for the various groups remains a highly subjective procedure for as long as the examiner has to make the final determination. This process that Roe describes is a clear illustration of the problems inherent in scoring Qs without some objective scoring standards. This problem has existed since the Rorschach came into being. It has not only failed to be resolved, it has almost literally failed to be addressed. Roe's discussion, quoted above, is the most direct and lengthy treatment of the subject that was found in reviewing the literature. This situation has been aptly summarized by Rickers-
Ovsiankina who observes that the Q scoring category "has failed to stimulate a notable amount of either research or theoretical speculation." (1960, p. 18)

Another problem that arises in dealing with the Q category concerns the average expectable range of Q responses. If an Q response is a once-in-a-hundred response and if the average number of responses per protocol is approximately 50, does it follow that the average expectable number of Q responses per protocol is one? That is both a reasonable and an unreasonable assumption. It is reasonable in the sense that the mathematics involved seem to work to a figure of 0.5 Q responses per record, although a moment's reflection makes it clear that this arithmetic-type reasoning can be, and probably is, specious in this instance. Dividing the number of responses per record into the quantitative scoring criterion is like mixing apples and oranges - the number of responses per record should be divided by the number of Q responses per same record; or alternatively, the 1:100 standard should be applied to the responses per specific blot area that are given across subjects. It makes no sense to arrive at a mean number of expectable Q responses per record by using the reasoning stated above. It does make sense, however, to question whether or not the reasoning customarily employed when dealing with the Q category is similarly skewed.

Obviously any number of Qs per protocol is possible, yet it is logical to expect that there is some meaningful range for such responses. Klopfer et al. (1954) state that from 25% to 50% of the total number of responses per record is an acceptable range. By this they mean that
such a range may be expected in an average record, but they present no evidence to support that statement. Hermann Rorschach (1942), too, attempted to present guidelines for numbers of Qs per record. He hypothesized that constricted, unimaginative subjects would produce as many as 40% Qs per record. Qs well in excess of 50% are considered indicative of psychotic processes. The net result of all this seems to be that in an average 20 to 45 response record, twice as many Q responses as other responses (i.e., 6 to 10 responses, 25 to 40%) constitute an acceptable range. This rough estimate is crudely substantiated by Roe's data. Her Q totals consistently show a mean Q percentage around 20% (1949b, 1952a), although in one article (1953) the group mean was approximately 32% (N = 22, Rs = 1473, Qs = 471). In another article she reports a range of Qs from 9% to 43% (1949a). In a similar manner the data of the pilot study for the present study indicate that 30% or more of all responses are Qs when college students are used as subjects and the Hertz tables are used for scoring. These figures are to be considered tentative at this point, however, and do not comprise substantive evidence, meaning that the issue of the acceptable or expectable number of Qs per Rorschach protocol remains unsettled. It also seems to be a question that has received little explicit attention in the literature.

Yet another problem with this rough-and-ready scoring standard is that it seems to require that Rorschach testers not use the Q category for at least the first 100 protocols that they administer and score. Furthermore, this overlaps with the problem concerning the testing population from which the tester has gathered his experience.
This seems to have some fairly serious implications. For example, it seems logical to conclude that students cannot be taught to score Q responses during their graduate school careers - unless they are required to administer over 100 Rorschachs during their training period. More importantly, there seems to be some implicit assumption that all persons who take the Rorschach test are psychometrically interchangeable. This seems to follow logically from the statement that a tester is qualified to score Q's once he has administered and scored 100 protocols. That seems to say that any 100 protocols will do, just so long as the scorer has the experience of administering and scoring 100 protocols. The actual nature of the subject sample is presumably immaterial. The Q category seems to be the only Rorschach scoring category where this applies. The other scoring categories involve definite, objective scoring guidelines so that any score given can be evaluated by a second scorer according to some externally verifiable feature of the response. Again, this issue has attracted little, if any, attention in the testing literature.

The second of the four truisms that appear in the Rorschach literature is that Q responses result from statistically uncommon meanings assigned to frequently seen blot areas. This conclusion follows from the definition of an Q response. Presumably the stimuli to which the subjects respond are fairly common, but the content with which they respond is uncommon. Said another way, what makes a response an Q response is the nature of the percept and not the nature of the stimulus. It is as if test subjects respond to essentially the same blot features and it is only when a given subject projects a
statistically infrequent percept onto the standard blot area that an $O$ response results. Experience indicates that many $O$ responses result from this process, but experience also indicates the presence of a totally different process of generating $O$s.

The latter process involves responding to an area of the blot that elicits responses from almost no one else. This seems to be a totally different method of producing $O$s than the method that has traditionally been described. The blot area chosen, rather than the associative content, seems to be the defining characteristic. When a subject responds to a truly rare blot feature, the actual content of the response is immaterial in the scoring. The content of the response may be prosaic or esoteric but in either event the response will be scored $O$.

Now this immediately presents another scoring problem because a way must be discovered for determining what is a common blot area and what is an original one. In other words, the same problem that exists regarding response content also exists regarding response stimulus. The tester in this instance would need to have not only a mental catalogue of response content but also of blot areas as response stimuli in order to determine what is or is not a 1:100 response. The testing literature contains little discussion of this particular topic, although Klopfer's definition of an $O$ response was one factor in raising the question while Phillips and Smith state "The less frequent a content, the more important to the subject are the materials which are revealed." (1953, p. 112) Statements such as these seem to suggest that the content of $O$ responses is of special dynamic importance and that testers ought
to focus primarily on the content of $Q$ responses. There is no discussion that the present investigator knows of in the testing literature that pertains specifically to the topic of the blot area(s) used in the production of Rorschach $Q$ responses.

This topic needs to be pursued further. The idea that response area may be as important as response content in the production of $Q$ responses is a reasonable idea that deserves careful consideration. Furthermore, it is reasonable to wonder why the subject has received no attention to date. It suggests that one way was originally proposed for viewing $Q$ data and that one way has not been challenged, expanded or in any way altered. It is quite surprising that the literature does not seem to contain any discussion of the factor of blot area - if there are truly countless numbers of response associations possible to a blot area, there are nearly countless ways to divide up a blot area so that in effect there are almost unlimited blot areas to which subjects may respond. What may be original about a response is not the associative content but rather the unique area of the blot that was utilized.

The third of the four truisms that exist in the Rorschach literature is that the scoring method for $Q$s relies entirely on the tester's recall of his past testing experience. This subject has already been touched upon in the discussion of the two prior truisms stated above, but it is actually a separate issue worthy of separate discussion. The entire topic is suggested by H. Rorschach's original suggestion of the 1:100 criterion and Klopfer's subsequent statement that $Q$ responses are "those that occur as rarely as once in one
hundred records in the experience of the individual examiner."
(1954, p. 22) This definition puts the entire burden of scoring upon the
tester's recall for recent and past events. This is clearly an
unreliable procedure and it is difficult to believe that there has been
no direct discussion of this issue in the literature on projective
techniques, especially the voluminous Rorschach literature.

The issue of the reliability of an examiner's memory would be
largely obviated if there were frequency tables of Rorschach responses
broken down into scoring per response per blot area for each of a
standardized number of blot areas. Of course, such a procedure would
present a whole set of questions and procedural problems but it would
seem to move the difficulties to a more advanced level. One of the
obvious questions involved in this would be what constitutes an
acceptable subject sample on which to compile the frequency table.
This question would ordinarily be beyond the scope of a study like this
but it has relevance here because of its methodological implications.
There are a number of frequency tables of Rorschach responses in the
current literature. Small (1956), Thomas, Ross, and Freed (1964, 1965),
and Hertz (1951) have all published such tables. These are the only
frequency tables, known to this investigator, that are published for
the specific purpose of reporting the responses of large numbers of
subjects for the use of Rorschach testers.

The Small tables are not appropriate for use in scoring 0
responses specifically; the tables published by Thomas et al., report
only the responses of medical students; and the Hertz tables report only
the responses of 11 to 16-year old Cleveland school children, "all of
approximately high average intelligence." (1951, p. 3) None of these claim to report the Rorschach responses of normal adult subjects and this highlights the lack of attention that has been given to the subject of scoring Qₜ responses. Since these tables are all that exist at present, however, one of them ought to have value in helping to fashion some objective standard to be used in scoring Qₜs. There is a need to empirically determine what response criterion meets the statistical criterion for originality, just as there is a need to know what blot areas are responded to less-than-one-in-one-hundred-times per card.

The fourth of the four truisms that exist in the Rorschach literature is that persons who produce relatively high numbers of Qₜ responses demonstrate manifest originality or resourcefulness in their everyday lives. This assumption was originally made by Rorschach himself and it has since been passed along, untested and unproven, from one generation of test users to another. The series of studies by Roe, cited above, attempted to generate data pertinent to this assumption. Her subjects were physical scientists (biologists and physicists) from eight universities plus the National Museum, and psychologists from seven universities, and anthropologists from five universities. All subjects were specifically selected because they had demonstrated high degrees of intelligence, resourcefulness, originality and achievement in their work. These are the very kinds of persons who would be expected to give high numbers of Qₜs, as a group.

By and large, Roe's findings tend to dispute - or at least they fail to support - the previously untested assumptions regarding
Q responses. She reported that her 382 subjects gave a total number
of responses of 13,675 with a mean number of responses per record of
35.8 and a mean number of P responses of 5.64 and a mean number of
Q responses of 6.82. Both the P and the Q responses were more frequent
than the responses in any other content category except A (1952b).
In other words, Q responses were the second highest category of
responses in the entire test sample. This suggests that there are a
large number of possible responses capable of meeting the 1:100
criterion, and that such responses can occur with considerable
frequency in most records. In one of the studies Roe (1952a) reported
that the mean Q percentage for the physical scientists as a group is
18.6%; the same measure for biologists as a group is 21.5% (1949b);
and in a separate study (1953), she reported a total Q percentage of
30.6% for the group of psychologists (total number of responses = 924,
and total number of Qs = 283) and a total Q percentage of 34.3 for
anthropologists (total responses = 549 and total Q = 188).

These findings do not support the claim that high numbers of
Q responses on a Rorschach protocol reflect manifest originality on
the part of the test subjects. These data provide no support for the
assumption that highly original scientists give significantly greater
numbers of Q responses than their apparently less original peers.
Indeed, this was one of the disappointments of Roe's series of studies.
It was expected that her research would provide some increased under-
standing of the Q scoring category and to the degree that this occurred
the knowledge was of a negative nature. Her findings tended to
contradict rather than confirm the untested clinical assumptions about
0 responses. Following this, interest in the topic of 0 responses seemed to decline to the point where no further research has been reported. The testing texts tend to retain the 0 scoring category but to treat it briefly, while the research literature contains virtually no mention of the subject. The literature seems to say, in effect, that the 0 scoring category is an honored anachronism - it is not really understood, but its background makes it too revered to discard.

Research on the Embedded Figures Test: The Embedded Figures Test (EFT) is a test that reliably measures a person's ability to locate a certain designated perceptual figure within a larger and more complex visual whole. The test originally consisted of 24 pairs of stimulus cards. One card was a relatively simple visual figure, and the other card was a more complex figure. The cards were presented one at a time to the subject who was instructed to find the simple figure in the complex one. This test was used extensively in spatial orientation research (cf., Witkin, Lewis, Hertzman, Machover, Hiessner, & Wapner, 1954) and was found to be a highly reliable instrument for measuring specific kinds of perceptual functioning. For instance, Dana and Goocher (1959) report that for a sample of 25 subjects - 17 female (mean age 24.7) and 8 males (mean age 27) - the test-retest Pearson product-moment reliability coefficients were: .92 for mean time to locate the simple figures; .61 for number of correct solutions; and .87 for number of correct solutions per unit of time. These authors concluded, "Subjects tended to maintain the same relative position from test to retest. T-tests for mean time per solution and number of
correct responses on test-retest over a 1-week interval ($t = 7.4$ and $4.2$, respectively), were significant at less that the .001 levels." (1959, pp. 100-101) Witkin reports similar findings - odd-even reliability coefficients of .87 for 51 males, and .74 for 51 females - and he also states "women, on the average require considerably more time to detect the simple figure than do men." (1950, p. 15)

Corrected odd-even correlations of .88 (Loeff, 1961) and .95 (Gardner, Jackson & Messick, 1960) have also been reported. Leona Tyler, in the Mental Measurements Yearbook, summarizes the findings of many studies by stating, "Test-retest coefficients for men and women, even with a three-year interval between administrations, were .89. Stability coefficients over shorter intervals and split-half coefficients have tended to run even higher." (1965, p. 212)

Because of these high reliabilities a reduction of the length of the test seemed possible. Jackson (1956) conducted an item analysis of all 24 test items and recommended the use of a 12-item test and reduction of the time limit per trial from five minutes to three minutes. Witkin, Dyk, Paterson, Goodenough, and Karp support Jackson's suggestion, stating "Our own data indicate that the first twelve items in the standard order of presentation provide about as reliable a measure as any twelve items drawn from the total set." (1962, p.40)

It has become accepted procedure to use the first twelve items of the EFT for research purposes.

The EFT has become a frequently-used, well-researched instrument. Witkin and his colleagues popularized the test in the 1950's by using it as part of a large research study that led to many publications.
The study revealed what appear to be two distinct modes of perceiving, termed field-dependent and field-independent. The former mode is characterized by a relative inability to see stimuli independently of the visual context in which they occur, while the latter mode is characterized by a relatively high ability to separate a target stimulus from the surrounding perceptual field. Gough summarizes the situation - again in the *Mental Measurements Yearbook* - by stating "one of the most attractive features of this test is its firm anchoring in a systematic context of theory and empirical evidence." (1965, p. 210) The EFT is accepted as the single most practical means of measuring field dependence/independence.

The EFT is included in this study because of its value in measuring the two perceptual field-related modes. The field-independent mode seems to have some apparent similarity to one of the two processes that seem to produce Rorschach Qs responses. It will be remembered that there seem to be two distinct processes that produce Qs - one involves responding to uncommonly-seen blot areas with almost any response, and the other involves uncommon content responses to commonly-seen blot areas. For the sake of convenience the uncommon area/common content Qs will be termed Type One Qs, while the uncommon content-common area Qs will be termed Type Two Qs.

It seems reasonable to inquire whether or not there is some relationship between the field-independent perceptual mode and the process that leads to Type One Qs. Field-independent perceiving involves analyzing a complex visual configuration and eventually responding to some parts of it while ignoring the rest. This involves
breaking the whole into parts or focusing on a portion of a whole and
de-emphasizing the remainder. In responding to a portion of the whole,
that portion has to be mentally separated out from the whole. Since it
is the perceiver's product he or she is free to define any boundaries
for the percept. There can be as many different boundaries and percepts
in response to any given perceptual whole as there are perceivers.
This is what seems to happen in the production of Type One Os - a
person comes to a stimulus blot and carves out a portion of the whole.
If the examiner inquires closely for the particular area that defines
the percept it becomes apparent that many different blot areas are used
in the service of the same response content (for example, "witch" at
the top of Card IX, or "man with a beard" on the side of Card V, or
"clouds" on Cards III, IV, V, VII and IX). The important feature
here is that the exact limits of the percept be traced and recorded by
the examiner. When this is done many Type One Os appear, and, as a
rule, they seem to result from a perceptual process that appears to
conform quite closely to the field-independent mode. This is the
rationale for including the EFT in a study concerning the concept of
originality. An attempt will be made to determine whether or not a
consistent relationship exists between Type One O responses and
field-independent perceptual functioning.

There is an additional sense in which the EFT may prove useful
to the present research. The EFT has been used in studies having a
wide range of target behaviors - for instance it has been used to
measure field-dependence/independence in relation to such internal
psychological variables as concept formation (Elkind, Koegler and Go,
1963), intellectual functioning as measured by standard intelligence tests (Goodenough and Karp, 1961), achievement motivation (Wertheim and Mednick, 1953), need for approval (Cooper, 1964), fear of approval (Heckhausen, 1967), and self-esteem (Coopersmith, 1967); and such overt behavioral variables as asthmatic symptomatology (Fishbein, 1963), activity-passivity (League and Douglas, 1961), alcoholism (Karp, Poster, and Goodman, 1963) and obesity (Pardes and Karp, 1965). It is also reported that persons in whom paranoid symptoms are evident have been shown to be field-independent, while obsessional patients tend to be more field-independent than a comparable group of hysterical patients, who tended to be more field-dependent (references cited in Lewis, 1971, pp. 137 and 140).

Furthermore, the field-dependence/independence dimension has been shown to be related to a constellation of personality factors that are coming to have increasing theoretical significance. According to the first research by Witkin and his colleagues on the relation between perceptual modes and personality characteristics, field-independent perceivers exhibit "activity and independence in relation to the environment; closer communication with, and better control of, their own impulses; and ...relatively high self-esteem and a more differentiated, mature body image." (Witkin, et al., 1954, p. 469) Field-dependent perceivers, on the other hand, show tendencies toward "inability to function independently of environmental support, an absence of initiating activity, and a readiness to submit to forces of authority" (ibid, p. 467); and they also "tend to be characterized by passivity in dealing with the environment; by unfamiliarity with and fear of their own
impulses, together with poor control over them; ...and by the possession of a relatively primitive, undifferentiated body image." (ibid, p. 469)

Subsequent research with these two distinct sets of psychological characteristics has led to some apparent refinement of the factors involved. The grouping of personality features associated with field-independent perceiving have been subsumed under the concept of "psychological differentiation". This concept has been heavily researched by Witkin and his colleagues, who report that "extent of definition of self-concept, articulateness of body image, and method of impulse regulation formed an interrelated cluster which is apt to be considered in evaluating people as more differentiated or less differentiated." (Witkin, et al., 1962, p. 8) In general, field-independence and more differentiation correspond quite closely, as do field-dependence and relatively less differentiation. The point here is that the rough constellation of characteristics first identified by Witkin in the research of the late 1940's and early 1950's has evolved into a fairly stable syndrome of personality factors that are being used in present personality research. (cf., Lewis, 1971) The EFT is considered a highly suitable measure of field-independence and psychological differentiation. This aspect of the EFT's applicability has obvious possibilities for the present study. Any strong relationship between EFT scores and Rorschach Q performance would promise to expand the usefulness of the Rorschach test, in general, and the Q category, in particular.

Research on the Remote Associates Test: The RAT is a 30-item
test with a 40-minute time limit. Each item consists of three words which all have a relatively common associative link with a fourth word. For example, the stimulus might be:

rat  blue  cottage

and the correct response would be "cheese". It is assumed that all American-born or -raised persons would have sufficient familiarity with English language usage to be appropriate subjects for the test. The test was developed by S. Mednick at the Institute of Personality Assessment and Research at the University of California. In the Examiner's Manual (Mednick & Mednick, 1967) for the test, reliability data are reported for three of the normative groups used. In a sample of 215 male undergraduates the odd-even reliability using the Spearman-Brown formula was estimated at .91; in a sample of 288 female undergraduates the same formula produced an odd-even reliability estimate of .92; and for a third group of 71 undergraduates (gender unspecified) the same formula gave an odd-even reliability estimate of .86. The latter group also took alternate forms of the test and a correlation of .81 was obtained between the two forms.

The RAT was developed as a measure of the ability to think creatively. Manifestly creative persons throughout history have reported in autobiographic materials that much of their creative thinking consisted of combining two previously unrelated elements into a new, unique product. Following this lead, Mednick has defined the creative thinking process as "the forming of associative elements into new combinations which either meet specified requirements or are in some way useful." (Mednick, Mednick, & Mednick, 1964, p. 85) In constructing the RAT he attempted to measure
each subject's ability to form associative elements into new, but clearly specified, responses. Thus he satisfied his own dual requirements of forging new associations and satisfying pre-existing standards of acceptability.

In the context of the present study, however, the RAT is not being used as a measure of creativity. This study deals with originality, not creativity per se. The difference here hinges on the usefulness requirement that Mednick posited. Originality will be considered as distinct from creativity in that the latter has a usefulness dimension and the former need not have. To be original, an item merely has to be statistically infrequent. The RAT is being used here as a measure of Type Two processes, which in Rorschach performance results in responses where the associative content produced - as opposed to the area of the blot used - is statistically infrequent and therefore original according to the definition used here.

Type Two originality seems to involve the ability to encounter relatively common stimuli and to generate a relatively uncommon response. On the Rorschach this means responding to a frequently-seen blot area with a statistically infrequent content response. High RAT scores seem to reflect the same ability. In constructing the test Mednick selected stimulus words that all occurred 100 times or more per million words in written English according to the Thorndike-Lorge (1944) word count. This means that all the stimulus words are relatively common. The response words, however, all are low probability ($p < .04$) responses to the stimulus words, meaning that the responses are uncommon responses to common stimuli. It is in this sense that the mental functioning that
results in high RAT scores seem to correspond to the kind of processes that produce Type Two Qs. This is the rationale for including the RAT in a study concerning the concept of originality. An attempt will be made to determine whether or not a consistent relationship exists between Type Two Q responses and performance on the RAT.

It should be mentioned that RAT scores have been shown to correlate with a number of variables other than the direct (i.e., creative job output) and indirect (i.e., rating scale scores, job grade classifications) measure of creativity reported in the Examiner's Manual. For instance, Houston and Mednick (1963) have shown that high-RAT scorers seek novelty more strongly than low-RAT scorers. Mednick (1962) has reported that high-RAT scorers also exhibited a general tendency to adopt more liberal attitudes than low-RAT scorers regarding personal beliefs and interpersonal relations. In a study by Higgins (1966) high-RAT scorers demonstrated more original problem-solving capability than low-RAT scorers when the problem-solving measures were number of original anagram solutions and total number of solutions. This latter finding tends to overlap with the finding of a separate study by Mendelsohn and Griswold (1964) who reported that high-RAT scorers utilize incidental stimuli better for problem-solving.

Such findings suggest that performance on the RAT correlates with a number of personality variables, some of which have already been researched and reported in the literature. In this sense, RAT scores are suggestive of a complex of personality factors in a manner similar to the relationship between EFT scores and the differentiation syndrome reported earlier. Any relation between RAT scores and Rorschach Q scores
would therefore have implications for defining or understanding the clinical significance of \( Q \) responses.
CHAPTER III

METHODOLOGY

Subjects

This study uses 40 subjects, 20 males and 20 females, most of whom are college age. Almost all subjects are Loyola undergraduate students; but because some undergraduates are considerably older than others, an attempt was made to obtain some subjects who were younger than college age. The age range for the total sample is 17 to 33 years. For males the range is 17-31 years old, with a mean age of 21.2 years and a standard deviation of 4.06. For females the range is 17-33 years old, with a mean age of 20.9 and a standard deviation of 8.40.

College age subjects are used for a number of reasons. The most obvious one is availability. There was no evident reason for believing that any particular group of subjects would give more appropriate data than any other group, and so availability became a primary consideration in obtaining subjects. A second major consideration was that the Hertz frequency tables were used and those tables were standardized on adolescents. It was felt that it was desirable to use subjects at the younger end of the adult range in order to more nearly approximate the age of the criterion sample while still dealing with adult responses. Finally, undergraduates were used because such a sample makes replication easier, and much of the methodology of this study was adopted with the idea of replication in
Examiners

There were a total of 8 examiners. The writer was senior examiner. In addition to the senior examiner there were 7 student examiners. The student examiners were all psychology graduate students enrolled in a two-semester course on personality testing. As part of their class assignment they were to administer 8 projective batteries, which included Rorschachs. The class was asked to volunteer to be involved in the present study, and all who volunteered were accepted as student examiners and were trained in the techniques to be used. First the purpose of the study was explained and then the special Testing of the Limits (TOL) procedure which was developed for this study was explained to the student examiners. They were instructed as a group in the administration of the TOL. In addition, written instructions were included in the packet of materials to be used in each TOL administration. (See Appendix A for all TOL materials.) The student testers were instructed to administer the TOL to all subjects that they tested with the Rorschach. When they scheduled a Rorschach administration they notified the senior examiner, who observed their administrations of the TOL.

One TOL administration was rejected because of a failure by the examiner to restrict herself to the written instructions. She was testing a personal friend and it appeared that she began prompting responses during the TOL. Nothing was said to the tester during the administration but when the testing session was ended this topic was raised and discussed, and the data she collected was discarded.
Twenty subjects were tested by the student examiners. Twenty subjects were tested by the senior examiner. All subjects tested by the student examiners were subsequently administered the EFT and RAT by the senior examiner, who also administered the EFT and RAT to all the subjects he tested.

Tests and Procedures Used

Three psychometric tests and one new TOL procedure were used in gathering the data. In addition, one validation check and a series of rank-order correlation coefficients were used to evaluate the data. The psychometric tests were: The Rorschach test, the Embedded Figures Test, and the Remote Associates Test. The TOL procedure was devised specifically for this study.

The purpose of the Rorschach test was to generate a baseline figure for number of responses. The Rorschach was administered according to standard instructions. The protocols so obtained were scored for responses according to the instructions and the responses listed in the Hertz tables. Two persons scored each protocol for responses, and the scorings were correlated using a Pearson product-moment correlation coefficient. Differences in scoring were discussed by the two judges and resolved in all the baseline totals used.

EFT data take the form of time-to-completion scores. There are 12 trials in which the subject has the task of perceiving a stimulus figure within a larger perceptual gestalt. The time in seconds that each subject takes to locate the embedded figure is his/her score. Men consistently take less time on this task and because of that fact
it was decided that male and female data would be maintained separately. Also, the raw EFT results would have given data in which the lowest figures were the highest scores (i.e., best performance). This would have meant that the EFT data were in the opposite direction of the O and the RAT data. In order to correct for this, the results of each trial on the EFT were subtracted from 300, which is the maximum number of seconds allowable per trial. This brought the EFT figures in the same direction as the O and the RAT totals. The total EFT figure became a measure of how much of the allowable maximum time was not used, so that a high EFT score indicates high idiomorphic functioning.

RAT data take the form of number of correct responses to 30 test items within a 40 minute time limit. The response to each item is given a right-wrong scoring. The absolute number of correct responses is the test score. Good performance is indicated by high absolute values. This is consistent with the Rorschach and the EFT data. The RAT data are used here as a measure of idiographic functioning.

The TOL procedure is intended to elicit O responses that can be added to the baseline O data obtained during the standard Rorschach administration. It is possible that the standard Rorschach administration would produce a fair range of O's but it was considered desirable to include a testing procedure that would increase the number of O's. This was intended to ensure sufficient variability in O scores to satisfy assumptions underlying the use of the statistical procedures to be used. The TOL was constructed by inspecting every Hertz table for all Rorschach cards. The 12 areas that produced the
highest percentage of responses per total number of responses comprised the TOL areas. Three responses reported by Hertz for each area were selected. These three responses are termed empirical responses.

Two more responses were invented by the investigator and used with the three empirical responses. These two responses - termed nonsense responses - are never given to those areas, according to the Hertz tables. Those responses were arrived at by loosely associating to each of the blot areas until a response was produced that had little obvious correspondence to the blot area and that could not be found in the Hertz responses. The two nonsense responses bracket the first empirical response in each instance. Thus, there are five responses per blot area in the TOL and they occur in the sequence; nonsense-empirical-nonsense-empirical-empirical. This sequence was designed specifically to counteract a phenomenon reported by Huberman (1965), who observed that subjects tend to adopt a response set of accepting or rejecting all concepts suggested by test administrators during the standard Rorschach TOL. This TOL for responses is a new procedure, unique to this study, and the content of the TOL is also new.

A final procedure is one which results in a "Total 0" category and reflects the sum of the two separate procedures used to elicit or generate responses on the Rorschach. The number of baseline responses for each protocol were obtained by adding all the main responses (obtained during the Free Association phase) and one-half of all the additional responses (obtained during the Inquiry phase or during the TOL for Originals). In the TOL the number of empirical responses that received an affirmative reply were subtracted from the number of nonsense responses that received
an affirmative reply, and then that figure was halved. The number of Os obtained from the standard Rorschach administration and the number obtained from the TOL were recorded separately. Separate headings were maintained for the number of Os given spontaneously (#O), the number obtained from the TOL procedure (TOL), and the total number generated by these testing conditions (TOTAL 0). The purpose of this was threefold. First, to determine what percentage of the total number of responses on the Rorschach can be expected to be Os responses, using the Hertz tables in scoring. Second, to determine whether the TOL does anything other than proportionately increase the number of Os responses that are given spontaneously (the use of this TOL procedure implies a belief that each subject's relative rank in terms of number of responses before and after the TOL will be essentially unchanged.) Third, to determine whether or not there is a difference between using only spontaneous Os in the statistical procedures and using the combined Os. These issues all have direct relevance for considerations of methodology to be used in any future research with Os responses.

Tabulations and Statistics:

A Pearson product-moment correlation coefficient was computed to measure interscorer reliability of the baseline Os scorings. These data were presented in Tables 1 and 2.

The various Os measures and the EFT and RAT data for each subject are presented in Tables 3 and 4.

Rank-order correlation coefficients were computed to determine the degree to which the following measures co-varied uniformly: number of Os per Rorschach protocol; number of total responses per Rorschach
TABLE 1

SCORING OF MALE Q DATA BY TWO JUDGES
USING HERTZ TABLES

<table>
<thead>
<tr>
<th>Subject</th>
<th>Judge #1</th>
<th>Judge #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>JS</td>
<td>26.5</td>
<td>21.0</td>
</tr>
<tr>
<td>RW</td>
<td>13.0</td>
<td>10.0</td>
</tr>
<tr>
<td>GG</td>
<td>12.0</td>
<td>10.0</td>
</tr>
<tr>
<td>JA</td>
<td>7.0</td>
<td>6.0</td>
</tr>
<tr>
<td>RB</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>RD</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>DP</td>
<td>6.0</td>
<td>5.0</td>
</tr>
<tr>
<td>TJ</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>EK</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>MM</td>
<td>9.0</td>
<td>9.0</td>
</tr>
<tr>
<td>BG</td>
<td>4.0</td>
<td>3.0</td>
</tr>
<tr>
<td>DP</td>
<td>38.0</td>
<td>31.0</td>
</tr>
<tr>
<td>JH</td>
<td>7.0</td>
<td>6.0</td>
</tr>
<tr>
<td>RD</td>
<td>9.0</td>
<td>8.0</td>
</tr>
<tr>
<td>TB</td>
<td>23.0</td>
<td>19.0</td>
</tr>
<tr>
<td>JC</td>
<td>5.0</td>
<td>7.0</td>
</tr>
<tr>
<td>RM</td>
<td>16.0</td>
<td>13.0</td>
</tr>
<tr>
<td>FA</td>
<td>7.0</td>
<td>5.0</td>
</tr>
<tr>
<td>JF</td>
<td>10.0</td>
<td>8.0</td>
</tr>
<tr>
<td>PIO</td>
<td>27.0</td>
<td>20.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge #1</td>
<td>233.5</td>
<td>11.7</td>
</tr>
<tr>
<td>Judge #2</td>
<td>195.0</td>
<td>9.8</td>
</tr>
</tbody>
</table>

Pearson r = .984

P ≤ .001
TABLE 2

SCORING OF FEMALE Q DATA BY TWO JUDGES
USING HERTZ TABLES

<table>
<thead>
<tr>
<th>Subject</th>
<th>Judge #1</th>
<th>Judge #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>18.0</td>
<td>14.0</td>
</tr>
<tr>
<td>AE</td>
<td>37.0</td>
<td>34.0</td>
</tr>
<tr>
<td>FL</td>
<td>31.0</td>
<td>32.0</td>
</tr>
<tr>
<td>RK</td>
<td>9.0</td>
<td>4.0</td>
</tr>
<tr>
<td>DY</td>
<td>3.0</td>
<td>1.0</td>
</tr>
<tr>
<td>CD</td>
<td>21.0</td>
<td>16.0</td>
</tr>
<tr>
<td>VD</td>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>SS</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CB</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>JK</td>
<td>23.0</td>
<td>15.0</td>
</tr>
<tr>
<td>LH</td>
<td>30.0</td>
<td>26.0</td>
</tr>
<tr>
<td>BS</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>XX</td>
<td>14.0</td>
<td>11.0</td>
</tr>
<tr>
<td>KR</td>
<td>13.0</td>
<td>9.0</td>
</tr>
<tr>
<td>ME</td>
<td>13.0</td>
<td>8.0</td>
</tr>
<tr>
<td>MW</td>
<td>59.0</td>
<td>48.0</td>
</tr>
<tr>
<td>AF</td>
<td>41.0</td>
<td>36.0</td>
</tr>
<tr>
<td>AC</td>
<td>31.0</td>
<td>25.0</td>
</tr>
<tr>
<td>SS</td>
<td>12.5</td>
<td>11.0</td>
</tr>
<tr>
<td>MF</td>
<td>22.0</td>
<td>18.0</td>
</tr>
</tbody>
</table>

Total = 414.5  345.0
Mean = 20.7  17.2

Pearson r = .982

P ≤ .001
### TABLE 3

**TOTAL TEST SCORES OF MALE SUBJECTS (N = 20)**

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>#R</th>
<th>#O</th>
<th>%O</th>
<th>TOL</th>
<th>TOTAL O</th>
<th>RAT</th>
<th>EFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD</td>
<td>22</td>
<td>6.0</td>
<td>42.9</td>
<td>26</td>
<td>19.0</td>
<td>29</td>
<td>3368</td>
</tr>
<tr>
<td>DP</td>
<td>17</td>
<td>6.0</td>
<td>35.3</td>
<td>16</td>
<td>11.0</td>
<td>18</td>
<td>3306</td>
</tr>
<tr>
<td>TJ</td>
<td>8</td>
<td>2.0</td>
<td>25.0</td>
<td>21</td>
<td>12.5</td>
<td>18</td>
<td>2998</td>
</tr>
<tr>
<td>EK</td>
<td>14</td>
<td>3.0</td>
<td>21.4</td>
<td>28</td>
<td>17.0</td>
<td>20</td>
<td>3297</td>
</tr>
<tr>
<td>DP</td>
<td>38</td>
<td>23.0</td>
<td>60.5</td>
<td>23</td>
<td>34.5</td>
<td>17</td>
<td>3380</td>
</tr>
<tr>
<td>RB</td>
<td>15</td>
<td>23.0</td>
<td>51.5</td>
<td>27</td>
<td>36.5</td>
<td>18</td>
<td>3256</td>
</tr>
<tr>
<td>GG</td>
<td>29</td>
<td>12.0</td>
<td>41.4</td>
<td>33</td>
<td>27.5</td>
<td>20</td>
<td>3231</td>
</tr>
<tr>
<td>JF</td>
<td>26</td>
<td>10.0</td>
<td>38.4</td>
<td>23</td>
<td>21.5</td>
<td>17</td>
<td>3297</td>
</tr>
<tr>
<td>PTO</td>
<td>51</td>
<td>27.0</td>
<td>52.9</td>
<td>23</td>
<td>38.5</td>
<td>13</td>
<td>2713</td>
</tr>
<tr>
<td>RW</td>
<td>20</td>
<td>13.0</td>
<td>65.0</td>
<td>16</td>
<td>21.0</td>
<td>5</td>
<td>3092</td>
</tr>
<tr>
<td>MM</td>
<td>24</td>
<td>9.0</td>
<td>37.5</td>
<td>21</td>
<td>19.5</td>
<td>1</td>
<td>2267</td>
</tr>
<tr>
<td>RG</td>
<td>15</td>
<td>4.0</td>
<td>26.7</td>
<td>10</td>
<td>9.0</td>
<td>20</td>
<td>3214</td>
</tr>
<tr>
<td>JS</td>
<td>64</td>
<td>26.5</td>
<td>41.6</td>
<td>33</td>
<td>13.0</td>
<td>19</td>
<td>3078</td>
</tr>
<tr>
<td>JH</td>
<td>17</td>
<td>7.0</td>
<td>41.2</td>
<td>19</td>
<td>16.5</td>
<td>2</td>
<td>1386</td>
</tr>
<tr>
<td>RD</td>
<td>26</td>
<td>9.0</td>
<td>34.6</td>
<td>30</td>
<td>24.0</td>
<td>6</td>
<td>3079</td>
</tr>
<tr>
<td>JA</td>
<td>18</td>
<td>3.0</td>
<td>16.7</td>
<td>19</td>
<td>12.5</td>
<td>10</td>
<td>2582</td>
</tr>
<tr>
<td>JC</td>
<td>26</td>
<td>5.0</td>
<td>19.2</td>
<td>15</td>
<td>12.5</td>
<td>18</td>
<td>3194</td>
</tr>
<tr>
<td>FA</td>
<td>14</td>
<td>7.0</td>
<td>50.0</td>
<td>29</td>
<td>21.5</td>
<td>13</td>
<td>3105</td>
</tr>
<tr>
<td>RM</td>
<td>38</td>
<td>16.0</td>
<td>42.1</td>
<td>17</td>
<td>24.5</td>
<td>19</td>
<td>3423</td>
</tr>
<tr>
<td>RB</td>
<td>12</td>
<td>3.0</td>
<td>25.0</td>
<td>22</td>
<td>14.0</td>
<td>18</td>
<td>2120</td>
</tr>
</tbody>
</table>

**TOTAL**  524  214.5  768.7  451  439.0  301  59713

**MEAN**  26.2  10.7  38.4  22.5  21.9  15.0  2985.6
TABLE 4

TOTAL TEST SCORES OF FEMALE SUBJECTS (N = 20)

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>#R</th>
<th>#O</th>
<th>%O</th>
<th>TOL</th>
<th>TOTAL 0</th>
<th>RAT</th>
<th>EFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RK</td>
<td>30</td>
<td>9.0</td>
<td>29.7</td>
<td>25</td>
<td>21.5</td>
<td>19</td>
<td>2675</td>
</tr>
<tr>
<td>Dv</td>
<td>11</td>
<td>3.0</td>
<td>27.3</td>
<td>20</td>
<td>13.0</td>
<td>12</td>
<td>2280</td>
</tr>
<tr>
<td>VD</td>
<td>53</td>
<td>20.0</td>
<td>37.7</td>
<td>27</td>
<td>33.5</td>
<td>20</td>
<td>2858</td>
</tr>
<tr>
<td>SS</td>
<td>16</td>
<td>1.0</td>
<td>25.0</td>
<td>16</td>
<td>12.0</td>
<td>15</td>
<td>2067</td>
</tr>
<tr>
<td>CB</td>
<td>16</td>
<td>5.0</td>
<td>31.3</td>
<td>17</td>
<td>13.5</td>
<td>9</td>
<td>2776</td>
</tr>
<tr>
<td>BS</td>
<td>23</td>
<td>8.0</td>
<td>34.8</td>
<td>28</td>
<td>22.0</td>
<td>19</td>
<td>2765</td>
</tr>
<tr>
<td>KR</td>
<td>36</td>
<td>13.0</td>
<td>36.1</td>
<td>25</td>
<td>25.5</td>
<td>15</td>
<td>3219</td>
</tr>
<tr>
<td>FL</td>
<td>62</td>
<td>31.0</td>
<td>50.0</td>
<td>30</td>
<td>46.0</td>
<td>8</td>
<td>3114</td>
</tr>
<tr>
<td>AE</td>
<td>47</td>
<td>37.0</td>
<td>78.7</td>
<td>23</td>
<td>48.5</td>
<td>13</td>
<td>1473</td>
</tr>
<tr>
<td>ME</td>
<td>34</td>
<td>13.0</td>
<td>38.2</td>
<td>21</td>
<td>23.5</td>
<td>19</td>
<td>3008</td>
</tr>
<tr>
<td>LH</td>
<td>59</td>
<td>30.0</td>
<td>49.9</td>
<td>27</td>
<td>43.5</td>
<td>9</td>
<td>3056</td>
</tr>
<tr>
<td>JK</td>
<td>64</td>
<td>23.0</td>
<td>35.9</td>
<td>37</td>
<td>41.5</td>
<td>15</td>
<td>3093</td>
</tr>
<tr>
<td>DK</td>
<td>24</td>
<td>14.0</td>
<td>58.3</td>
<td>12</td>
<td>20.0</td>
<td>16</td>
<td>2947</td>
</tr>
<tr>
<td>AC</td>
<td>54</td>
<td>31.0</td>
<td>57.4</td>
<td>34</td>
<td>48.0</td>
<td>15</td>
<td>2788</td>
</tr>
<tr>
<td>SS</td>
<td>28</td>
<td>12.5</td>
<td>41.6</td>
<td>25</td>
<td>25.0</td>
<td>19</td>
<td>3033</td>
</tr>
<tr>
<td>MW</td>
<td>98</td>
<td>59.0</td>
<td>60.2</td>
<td>34</td>
<td>76.0</td>
<td>19</td>
<td>3114</td>
</tr>
<tr>
<td>AB</td>
<td>58</td>
<td>18.0</td>
<td>47.4</td>
<td>22</td>
<td>27.0</td>
<td>11</td>
<td>2702</td>
</tr>
<tr>
<td>CD</td>
<td>41</td>
<td>18.0</td>
<td>43.9</td>
<td>28</td>
<td>32.0</td>
<td>15</td>
<td>2769</td>
</tr>
<tr>
<td>MF</td>
<td>44</td>
<td>20.0</td>
<td>45.5</td>
<td>28</td>
<td>34.0</td>
<td>16</td>
<td>2831</td>
</tr>
<tr>
<td>AF</td>
<td>75.5</td>
<td>61.0</td>
<td>54.3</td>
<td>30</td>
<td>56.0</td>
<td>28</td>
<td>2877</td>
</tr>
</tbody>
</table>

TOTAL  853.5  409.5  886.2  509  664.0  315  55475
MEAN  42.7  20.5  44.3  25.4  33.2  15.7  2773.7
protocol; percent of Os per protocol; TOL Os; total Os per subject; EFT adjusted score; RAT raw score. A separate rank-order correlation coefficient (rho) was computed to test the relationship between each pair of variables. The rho coefficient was used because the assumption cannot be made that the O data will be normally distributed. A non-parametric statistic was used because the possibility of getting a restricted range of O data had to be respected. A separation between male and female data was maintained throughout in case there might be a sex difference in Rorschach O performance and because there is a known difference in EFT performance. A correlation matrix was constructed for males and for females (see Tables 5 and 6). Critical values were determined for all the rho coefficients so obtained.

An additional correlation matrix was constructed for combined male and female data, and critical values were determined for all the rho coefficients obtained (see Table 7). These data were combined in order to further investigate whatever relationships may exist but could not be predicted. Such a procedure seems consistent with the ground-breaking nature of this study - it is an attempt to present maximum amounts of data that may prove useful in generating future research.


<table>
<thead>
<tr>
<th></th>
<th>#O</th>
<th>#O</th>
<th>TOL</th>
<th>TOTAL</th>
<th>EFT</th>
<th>RAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>#R</td>
<td>.88***</td>
<td>.54**</td>
<td>.33</td>
<td>.82***</td>
<td>.14</td>
<td>.02</td>
</tr>
<tr>
<td>#O</td>
<td>.74***</td>
<td>.34</td>
<td>.92***</td>
<td>.16</td>
<td>-.01</td>
<td></td>
</tr>
<tr>
<td>#O</td>
<td>.26</td>
<td>.75***</td>
<td>.35</td>
<td>-.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOL</td>
<td>.67***</td>
<td>.12</td>
<td></td>
<td>.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td>.21</td>
<td>-.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFT</td>
<td></td>
<td></td>
<td></td>
<td>.48**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $P \leq .10$, two-tailed test
** $P \leq .05$, two-tailed test
*** $P \leq .01$, two-tailed test
### TABLE 6

**RANK-ORDER INTERCORRELATIONS BETWEEN FÖRSCHAVE MEASURES AND THE EMBEDDED FIGURES TEST (EFT) AND THE REMOTE ASSOCIATES TEST (RAT) FOR FEMALE SUBJECTS**

<table>
<thead>
<tr>
<th></th>
<th>#R</th>
<th>%O</th>
<th>TOL</th>
<th>TOTAL 0</th>
<th>EFT</th>
<th>RAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>#R</td>
<td>.96***</td>
<td>.62***</td>
<td>.80***</td>
<td>.95***</td>
<td>.52**</td>
<td>.12</td>
</tr>
<tr>
<td>%O</td>
<td>.83***</td>
<td>.67***</td>
<td>.97***</td>
<td></td>
<td>.35</td>
<td>.07</td>
</tr>
<tr>
<td>%D</td>
<td></td>
<td>.33</td>
<td>.76***</td>
<td>.77***</td>
<td>.11</td>
<td>.22</td>
</tr>
<tr>
<td>TOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.40</td>
<td>.06</td>
</tr>
<tr>
<td>TOTAL 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.14</td>
</tr>
</tbody>
</table>

*P ≤ .10, two-tailed test  
**P ≤ .05, two-tailed test  
***P ≤ .01, two-tailed test
TABLE 7

RANK-ORDER INTERCORRELATIONS BETWEEN RORSCHACH MEASURES AND THE EMBEDDED FIGURES TEST (EFT) AND THE REMOTE ASSOCIATES TEST (RAT) FOR ALL SUBJECTS

<table>
<thead>
<tr>
<th></th>
<th>#0</th>
<th>#0</th>
<th>TOL</th>
<th>TOTAL 0</th>
<th>EFT</th>
<th>RAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>#R</td>
<td>.94***</td>
<td>.60***</td>
<td>.60***</td>
<td>.84***</td>
<td>.07</td>
<td>.07</td>
</tr>
<tr>
<td>#0</td>
<td></td>
<td>.80***</td>
<td>.50***</td>
<td>.92***</td>
<td>.08</td>
<td>-.02</td>
</tr>
<tr>
<td>#0</td>
<td></td>
<td></td>
<td>.31</td>
<td>.76***</td>
<td>.17</td>
<td>-.03</td>
</tr>
<tr>
<td>TOL</td>
<td></td>
<td></td>
<td></td>
<td>.66***</td>
<td>.17</td>
<td>.17</td>
</tr>
<tr>
<td>TOTAL 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.07</td>
<td>-.03</td>
</tr>
<tr>
<td>EFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.42**</td>
</tr>
</tbody>
</table>

* P ≤ .10, two-tailed test
** P ≤ .05, two-tailed test
*** P ≤ .01, two-tailed test
CHAPTER IV

RESULTS AND DISCUSSION

Results

The data provide mixed support for the first hypothesis. The data for the male subjects reveal no significant uniform relationship between the EFT deviation scores and the Rorschach Q data. The data for the female subjects produce a significant correlation between EFT scores and the total number of responses to the standard administration of the Rorschach test (significant at the .05 level); a correlation between EFT data and the number of Q responses accepted during the Testing of the Limits phase that is significant at the .10 level; and a correlation between EFT scores and the total number of Q responses generated by the entire experimental procedure that is also significant at the .10 level. The correlations for female data between EFT scores and two other Rorschach variables are nonsignificant, although they approach significance at the .10 level. The combined male and female data are totally nonsignificant for all possible relationships between EFT scores and Rorschach measures. It is concluded that the data support the first hypothesis for female subjects but not for males. More will be said about this in the Discussion section.

The data do not support the second hypothesis. There is no significant relationship between the RAT scores and any of the Q variables. This lack of relationship holds for male and female data.
separately as well as combined. The correlation coefficients cluster around 0.00 with a range of .22 to -.15.

The third hypothesis is rejected although there is support in the data for the hypothesis stated. That is, there is no significant relationship between EFT deviation scores and the number of correct RAT scores for females, but there is a uniform relationship between the EFT scores and the RAT scores for males that is significant at the .05 level. There is also a uniform relationship between EFT scores and RAT scores for the combined male and female data that is significant at the .05 level. Thus two of the three tests of this hypothesis lead to rejection. It is concluded, therefore, that there is a significant uniform relationship between the EFT and RAT scores of male subjects.

The Pearson product-moment coefficients reveal interscorer reliability between the two scorers of the Q data of .984 and .982 for males and females, respectively. Both of these values are highly significant and indicate an extremely high degree of agreement between two independent scorers using the Hertz tables. This suggests that a standardized and reliable method of scoring for Q responses could readily be designed for future testing and research use, based on the Hertz tables.

Data were collected for seven distinct sets of variables. They were: EFT deviation scores (EFT); number of correct RAT responses (RAT); total number of responses to the Rorschach test (#R); total number of Q responses during the Free Association and the Inquiry phases of the standard Rorschach administration (#O); the percentage of the total
number of responses that the number of Os comprise (%O); the corrected number of Os elicited during the Testing of the Limits (TOL); and the total number of O responses obtained from the entire experimental testing procedure (TOTAL O). Rank-order correlations were computed for every possible pairing of these variables.

There is a consistent pattern to the correlation coefficients so obtained. The TOTAL O variable shows a significant correlation with the other four Rorschach measures (i.e., #R, #O, %O, and TOL). These relationships are significant at the .01 level in every instance and they apply to male, female, and combined data. The %O variable shows a significant correlation with all the other Rorschach variables except the TOL data for males and for females (although there is a significant correlation for the combined data). These relationships are all significant at the .01 level except for that between %O and TOL for the combined data, which is significant at the .10 level. The #O variable shows a significant correlation with all four of the other Rorschach variables except for the TOL data for females. Once again all the significant relationships reach or exceed the .01 level and apply to the male, the female, and the combined data except for the one instance mentioned.

There are two Rorschach categories that show variations from the general pattern of consistent intercorrelations among the Rorschach data. The two exceptional variables are the #R and the TOL headings. The #R data show a significant correlation, at the .01 level, with all the other Rorschach measures when using the female and the combined data. For male subjects the data correlate significantly for only three
of the four other variables (TOL is the exception), and of those three only two (#0 and TOTAL 0) are significant at the .01 level while the third variable (%0) is significant at the .05 level. For the TOL data the situation is even more discrepant. For females the TOL data produce correlations significant at the .01 level with #R, #0 and TOTAL 0 data; and a nonsignificant correlation with %0 data. For males the TOL data produce a single significant correlation - with TOTAL 0, at the .01 level. The TOL data for males show a statistically non-significant relationship with three of the four other Rorschach measures. For the combined data, TOL scores correlate at the .01 level with all other Rorschach measures except %0, where the correlation is still significant at the .10 level.

There is a significant relationship, at the .05 level, between EFT scores and #R for females but except for that one instance there are no statistically significant relationships between any EFT scores and any Rorschach data.

There are no statistically significant relationships between the RAT scores and any of the Rorschach measures, for either male, female, or combined data.

Discussion

The test of the first hypothesis has yielded some challenging data. There are no significant relationships between the EFT scores and any Rorschach measures for male subjects; for female subjects one of the five correlations between EFT scores and the Rorschach measures are significant at the .05 level, two are significant at the .10 level and the remaining two approach significance at the .10 level; and the intercorrelations for the combined data are lower than those for either
the male or the female data. In terms of subject performance there is a more uniform relationship between EFT scores and productivity on the Rorschach measures by women than by men.

This finding immediately raises a number of fairly compelling questions. For instance, why is there a sex difference? Or, since women are known to be low scorers, compared to men, on the EFT but they were high scorers on the Q measures here, does that mean poor EFT performance correlates with high Q production? Or, alternatively, is there a sex difference in Q performance rather than EFT performance that explains these findings? Looking at all of the EFT data suggests a very plausible answer. The combined data yield intercorrelations that are lower than those for the separate male and female data. This seems to indicate that the male and female data essentially represent two distinct subject populations. The EFT data must be separated by sex and examined as if male and female performance represents two separate variables. This conclusion is heightened by examining the EFT rankings which reveal that the 20 female EFT scores occupy the first 30 rankings. These data are consistent with the literature.

However, the literature was known from the outset so that (1) the data were separated, and (2) a correction factor was employed. This means that whatever sex differences appear in these conclusions are not primarily attributable to the fact that women are known to perform relatively poorly on the EFT. If the known sex difference in EFT performance does not seem to be a major factor, the foremost alternative is that there might be a sex difference in Q performance. And in fact there does seem to be such a difference. For the five
Rorschach measures the mean values for male data average approximately 50% lower than the scores for female subjects. Females scored higher as a group in all the Rorschach categories. Once the EFT raw scores were converted into deviation scores the female subjects generally had high EFT scores and high scores on the Rorschach measures, which accounts for the significance in the obtained correlations. No comparable correction was applied to the Q data from male subjects, however, and it might be that none is possible. Male subjects tended to give few spontaneous Qs, to accept fewer suggested Qs in the TOL, and to give fewer responses per test protocol than did female subjects. Male test performance appeared to be tight, critical and literalistic compared to the Rorschach behavior of female subjects.

Table 6 shows that EFT performance correlated most highly (significant at .05) with #R, then with TOL and TOTAL Q (significant at .10), then with #O. #R, TOL and #O seem to reflect a subject's willingness to respond affirmatively to what may appear as borderline concepts at times. Subjects who scored higher on those measures were those who were relatively more able or willing to take a loose approach to Rorschach responding. Female subjects were observed to do this with greater frequency than the male subjects, and it is inferred that this is a major factor in explaining the significant intercorrelations between the EFT scores and some of the Rorschach data for female subjects.

This says nothing about any possible relationship between idiomorphic perceiving and Q production, though. The presence of the sex difference is puzzling in this regard. If EFT performance reflects idiomorphic processes and if females score lower on the EFT but higher
on the Q measures, such a finding presents an obvious difficulty in demonstrating a relationship between idiomorphic processes and Q production. It can either be the case that idiomorphic perceiving actually does underlie the production of Rorschach Q responses, but the EFT is not an accurate measure of idiomorphic activity; or it can be the case that what have been termed here idiomorphic processes have no close correspondence to the number of Q responses given per subject. No available evidence seems to favor the former alternative. Empirical observations during testing seemed to indicate that Type One perceiving produced the majority of the Q responses that were obtained.

The most reasonable conclusion seems to be that the EFT is not a satisfactory measure of Type One activity, and that idiomorphic perceiving is not functionally equivalent to field-independent perceiving. Such a conclusion is accepted with regret because any reliable connection between the EFT and the Rorschach would have exciting possibilities for research and theory. This is especially true because of the relationship between EFT scores and the cluster of personality variables subsumed under the heading of psychological differentiation.

The results for the second hypothesis indicate that no useful relationship was discovered between RAT performance and the production of Rorschach Q responses. Whether this also indicates that there is no essential relationship between idiographic processes and Rorschach Qs remains an open question, although it seems likely that this would be the case. The overwhelming majority of the Q responses scored for this study seemed to be Type One rather than Type Two Qs. Thus
case-by-case observation reveals no relationship between idiographic processing and number of Rorschach Q responses. This also leaves open the question of whether or not the RAT measures idiographic functioning, although the evidence seems to favor a negative conclusion. The available evidence seems to be twofold. First, the RAT proves to be a somewhat mysterious measuring instrument. It claims to measure creativity and perhaps it really does, but it seems to measure a number of other things as well. The other thing(s) that it measures as adequately as it measures creativity raise the question of whether or not it actually measures some other more basic quality (or qualities) that relates in some way to creativity. The reception that the test has received by the field as a whole suggests that there is some lack of clarity about what it does in fact measure. It seems most prudent to conclude that the RAT probably does not measure an ability to associate highly uncommon content to common visual stimuli. A second type of evidence is that there were very few Type Two Qs observed during the Rorschach testings. It seems unlikely that there are very many Q responses that are the sole result of idiographic processing. If idiographic processing were to be viewed as the major source of Rorschach Q responses there would probably be so few Q that the entire scoring category could be dropped with little loss or regret. Since idiographic activity occurs so rarely it seems doubtful that this is what RAT scores reflect, except perhaps accidentally or, at best, only partially. Once again the conclusion to reject the stated hypothesis is reached with regret because it denies the possibility of relating number of Q responses to any of the personality characteristics correlated with performance on the RAT.
The data bearing on Hypothesis Three are somewhat equivocal. The correlation coefficients for the relationship between RAT scores and all of the Rorschach variables for both sexes, separately and together, cluster around zero, while the relationship between the RAT scores and the EFT scores of the male subjects and of the total combined sample are statistically significant at the .05 level. For females the relationship between RAT scores and Rorschach measures and between RAT and EFT scores were all nonsignificant. Thus, the only instance where RAT scores showed a significant uniform relationship with anything was with the EFT scores for males (which was also reflected in the combined data). The fact that the RAT should correlate with only one of six other measures and then for one gender only is a curious finding and seems to require some explanation.

At first glance there appears to be nothing inherent in the nature of the tests to account for such a finding. Males are known to perform better on the EFT and by and large they did so here but this should not be a factor when the data is separated by sex and rank-ordered. The same reasoning applies to the RAT data where women apparently perform slightly better as a group. The most plausible explanation in terms of other data from this study is that the male subjects tended, as a general rule, to perform more uniformly on the more structured, more clearly task-oriented tests (such as the EFT and the RAT) than on the less structured task presented by the Rorschach test. This explanation addresses the seemingly dissonant findings that for females there is a mildly significant uniform relationship between EFT performance and three of the five Rorschach measures while for males
there is a significant uniform relationship between EFT and RAT performance.

The reliability data are straightforward and even expectable when some objective scoring standard is utilized. The scoring decisions for most responses are simple to make, with few judgment demands. The Hertz tables and the Hertz scoring instructions were used and these covered most situations rather adequately. There were only three general types of situations that required the use of independent judgment and could lead to the introduction of scorer variability into the scoring process. The first of these had to do with blot area. There were a considerable number of instances where the blot area that the subject outlined was not in substantial agreement with the area Hertz indicates as consistent with the content involved in the response. The problem in such a situation is determining how much deviation from Hertz' limits is permissible and how much constitutes the basis for an 0 scoring for that response. In this study the two scorers used pilot project protocols to gain scoring practice, and in the process this very issue was raised and discussed. Both scorers seemed to gain a shared, general sense of how little deviation was consistent with the Hertz scoring and how much deviation warranted an 0 score.

A second type of situation that required a substantial degree of scorer judgment involved responses where the content was similar to, but not the same as, the response listed in the Hertz tables. Such situations are fairly common and are not easily resolved by a statement of a broad, general principle - they must be dealt with on an instance-by-instance basis. For example, Hertz may list "head of a
poodle" as an F+ response for a given area but not list "head of a sheepdog", "head of a fox", or "head of a hippo". If the scorer is willing to generalize from the listed response he is likely to accept one or more of the other responses as F+ responses; but if he is chary about generalizing he will probably score any of the other responses as an O response. In the scoring instructions Hertz encourages prudent generalization but this cannot preclude the introduction of interscorer variability.

The third major type of scoring problem arises when the response given by a test subject comprises either some combination of responses listed by Hertz or else the subject's response is a portion of a response listed by Hertz. In essence, the scorer has to deal with fractional responses in these situations. This is an especially intriguing problem because it has so many variations. For example, the response "Indian medicine-men with beards and wearing hats with buffalo horns, standing back to back and holding something at arm's length" was given to area S & S of Card VI. Hertz lists "man, with hand stretched out, holding out something," and "person's face with beard," and "person's face with crown" as F+ responses for area S; and "persons back to back" as an F+ response for area S & S. The problem for the scorer in this situation is obvious. The scorer must decide whether the entire percept as given is to be scored F+ since each element separately is F+, or whether the combination of each of these common elements into a new gestalt makes an O+ scoring mandatory. A second and slightly different problem that comes under this heading arose on Card II. A subject gave the response "Two elephants holding
a Christmas tree" to combined areas 5 + 5 plus 12. "Elephants" is an F+ response to area 5 + 5 and "fir tree" is listed as an F+ response to area 12: but there is no listing for both those percepts as one combined whole, just as there is no listing for the combining of area 5 + 5 with area 12. Again, the scorer's problem is obvious. A final related problem is when a response is given to only one side of a symmetrical blot area. For instance, the response "elephant" in the example given above is an illustration of this problem. The response given is F+ but it only uses half of the listed normally-used blot area. In this study the scorers agreed to score instances of unique combinatory wholes as 0; to score unlisted blot areas as 0; and to not score responses to half of a symmetrical whole as 0. Obviously, other standards could be formulated in an effort to standardize scoring. Obviously, too, such guidelines do not completely eliminate the need for scorer judgments in certain situations.

The Rorschach measures appear to have produced some fairly straightforward results. The TOTAL 0 data show highly significant (p < .01) correlations with the data of the remaining four variables. This is not really surprising in any instance. In the case of the relationship between TOTAL 0 and #O, as well as #O and TOL, significant correlations are readily explained by the fact that the latter three headings are all subsumed under the former one. It is possible that there could fail to be a significant correlation with the TOTAL 0 score, but the presence of such a correlation is not surprising. It is slightly more noteworthy that there is a significant correlation between TOTAL 0 and #R. Upon reflection, however, such a finding
merely suggests that as a subject gives more responses he or she also tends to get a higher 0 score. This combination is further substantiated by the fact that there is a highly significant correlation for both male and female data between #0 and #R. These data seem to support the common-sense expectation that the total number of responses given to the Rorschach blots and the number of 0 responses produced by any given subject will vary uniformly per subject.

The very high correlations between certain variables (for example, between TOTAL 0 and #0 for both males and females, and between #R and #0 for both males and females) suggested that there were overlapping measures or basic duplication of some measures. With this in mind, the correlation matrix was inspected with an eye toward eliminating certain data headings, if possible. The TOTAL 0 heading seemed most readily expendable because it seemed to overlap most obviously with other measures - most notably the #0 heading. Once an overlap between TOTAL 0 and #0 became obvious, the #0 category also began to appear more and more superfluous. This left the #R, #0, and TOL headings from the original five Rorschach measures maintained throughout this study.

The TOL data showed significant correlations with four of the five Rorschach variables for female subjects, with one of the Rorschach headings for male subjects, and with all the Rorschach headings for the combined data. What this really means is that it correlates significantly with TOTAL 0 for both sexes, and with #0 and #R for females. Viewed in this way the TOL heading adds relatively little to the total data picture. A significant correlation with TOTAL 0 is expected because
TOL data are an integral part of the TOTAL O figures. A significant correlation with #R for female subjects does contribute something to an understanding of the data for it suggests that females who are high-responders on the standard Rorschach will accept more suggested Qs during the TOL than will females who are low-responders or will male subjects in general. This is consistent with the previous conclusion that women tend to be more accepting, as a general rule, of possible Rorschach responses, either when they are responding spontaneously or when the examiner suggests responses.

Men, on the other hand, apparently approach the Rorschach task with a more critical outlook. The male TOL data showed low correlations with all the other Rorschach data except TOTAL O, as discussed previously. Such a finding seems to indicate that men, as a whole, are low-responders on all Rorschach measures. This is corroborated by the fact that men gave an average of 16.5 fewer total responses than females (mean #R = 42.7 for females, 26.2 for males); an average of 9.8 fewer Q responses during the standard administration (mean #Q = 20.5 for females, 10.7 for males); and an average of 3 less TOL responses (mean TOL = 25.5 for females, 22.5 for males). On a percentage basis these figures mean that female subjects gave 39% more total responses, 48% more Q responses, and 12% more TOL responses than male subjects. The apparent fact that men are more hesitant and more critical responders evidently underlies the fact that the TOL data for male subjects adds little to this study. The TOL data for female subjects is more informative but most of the information that those data provide had already been revealed through other measures. Thus, the TOL measure
also seems nonessential.

At this point the %O, TOL, and TOTAL O measures have all been
deemed nonessential for research purposes. However, the %O measure
does seem to possess potential clinical utility. In clinical settings
#0 is the accepted measure of originality, but it should be obvious
that %O is a more sophisticated measure. It appears preferable to use
%O rather than #0 when scoring protocols for clinical purposes.

The %O, TOL, and TOTAL O measures comprise the majority of the
special Rorschach methodology introduced in this study. To dismiss
these three measures is to dismiss the special procedures designed
specifically for this study in order to elicit maximum numbers of O
responses. It was considered desirable to make every effort to
generate Os because it was assumed that most subjects would not
spontaneously produce large enough numbers of Os to work with
statistically. Perhaps it is time now to seriously question that
assumption. Perhaps one of the foremost conclusions of this study can
be that all persons who take the Rorschach test will produce fairly
large numbers of O responses (in this study, mean number per subject
given by 20 males was 10.7 and by females 20.5), provided that an
objective scoring standard (in this study, the Hertz tables) is used
and a stringent outlining of all blot areas is done by the examiners.
No additional testing procedures are needed to obtain substantial
numbers of O responses from the standard Rorschach administration (in
this study, 48% of responses given by females were scored O, and 41%
of the responses given by males were scored O) when examiners are
trained to outline percepts exactly as given by the testing subjects.
Obviously this presumes that many or most Q responses will be considered original on the basis of the specific blot area used. This assumes that most Q responses result from what has been termed idiomorphic perceiving in this study. Conversely, this assumes that idiographic perceiving is not a major producer of Rorschach Q responses. It seems very unlikely in view of the subject behavior witnessed as part of the testing for this study that many Q responses result from truly unique content being associated to relatively commonplace blot areas. This is actually a major conclusion since those projective tests that mention this subject at all clearly convey the belief that Q responses are usually Type Two Qs that reflect idiographic processing. The results of this study contradict that belief.

Further, this study suggests that the Rorschach Q category actually measures a totally separate and different process which has little obvious similarity to the one presumed to be operative. In other words, originality as measured by the Rorschach test has more to do with the precise portion of the stimulus materials used by the subject than with the content of the response given or with the associative activity used by the subject in formulating a response. The actual content associated to a blot area probably has much less to do with an Q scoring - and therefore with defining originality in terms of the Rorschach test - than the exact area of the blot used does. This would seem to put the subject of Q responses in a new perspective.

For one thing, it seems to make Q responses much less mysterious and even exotic. Q responses have been made to seem like rare objects - the product of an unusual and perhaps even a nearly occult
coming-together-in-time of the proper blot area and a corresponding mental event in the sensorium of an extraordinary gifted subject. An O response somehow seemed like a slightly awesome mental product - the result of superior mental activity that could occur only rarely and only with certain subjects. The results of this study tend to discredit such a view and to suggest a very different alternate view. This study indicates that a substantial percentage of the total number of responses to the Rorschach cards can be scored O. Such a view corresponds very well with common-sense expectations, for it could be anticipated that different people would associate common content to similar but slightly different blot areas.

Another way in which this study might put a new light on the subject of O responses is by pointing out that if an O response does not reflect the operation of the kinds of mental processes that have been assumed up to now, perhaps the O category could be or should be dropped from Rorschach testing. The reasoning behind such a statement is that if O responses are not what they have been considered to be, the O category as presently defined has no real use and ought to be discarded. This study suggests that O responses do not reflect uniqueness or originality in thinking - or at least not the type of originality where one person looks at something that many other persons have seen and he suddenly experiences some startlingly original perception. This seems consistent with a study reported by Barron (1955) who found that Rorschach O+ and achromatic inkblots as measures of originality showed less correlation with manifest behavioral originality than did six other measures in his testing battery.
Whatever it is that inkblots measure, it seems that it is not demonstrable originality in overt behavior. Or, in other words, it appears that inkblot measures are inferior to other measures of originality.

If the Q category is not what it has been presumed to be, then what is it? This study suggests that it is a category that reveals how often subjects define the outlines of their percepts in statistically uncommon ways. Another way of saying this is that the Q category measures the degree to which each person responds to different or unique blot areas while taking the Rorschach test. Is this a behavior that psychologists, or anyone else, would want to measure? Does this behavior have any psychological significance - and if so, what is it? It might be that such a category has no special usefulness and it ought to be dropped because of that. This is certainly an issue raised by the present study.

Perhaps Barron's study can shed some light on this question. Barron's measures permitted him to dichotomize his subject sample into two groups, a high originality and a low originality group. In comparing the two groups Barron found five general characteristics that seemed to distinguish one group from the other. Those characteristics that distinguished the high originality group were: a preference for complexity and some degree of apparent imbalance in phenomena; more independence in their judgments; more self-assurance, ascendance and self-confidence in dealing with others; they "are more complex psychodynamically and have greater personal scope" (1955, p. 482); and they tend to entertain ideas and impulses that are commonly taboo. This latter characteristic (and
perhaps the first one, too) may have appeared in this study. This may underlie the fact that high $O$ responders tend to give many responses to the test materials. They are evidently lenient in accepting possible responses - they do not readily reject borderline percepts. Where low $O$ responders seem generally unwilling to report or to accept responses that do not correspond quite literally to "safe" blot features, high $O$ responders suppress fewer of their own mental products. Also, many of the $O$ responses tend to be obviously more complex and involved than non-$O$s. These observations will have to be subjected to further research, but if they prove to have validity it would indicate that there may be some genuine clinical significance for the $O$ category as viewed in the present study.

Alternatively, the $O$ scoring category could remain as is, although it seems that at least the use of some objective scoring standard is indicated. Let us assume for a moment that the $O$ heading is maintained and that the use of Hertz' tables is made mandatory when scoring Rorschach protocols. In such an eventuality matters would stay essentially where they are now except that the scoring would be more standardized and probably more reliable. When scoring $O$, the Hertz tables would be used primarily for content. Differences in blot areas used in generating the same or similar response content across subjects would be largely overlooked. Such a result is possible but not recommended, in view of the data of this study.

It seems that it would be unwise to overlook some of the conclusions of this study. The conclusion that many more Rorschach $O$ responses result from subjects responding to unusual blot configurations
rather than responding with statistically uncommon associative content seems to have important implications for Rorschach testers in clinical settings. The conclusion that very few Q responses seem to result from idiographic processes while notable numbers of Qs seem to result from idiomorphic processes seems to have many implications for research. The conclusion that neither the EFT or the RAT has any significant relationship with the number of Q responses given to the standard Rorschach, and furthermore, that neither test appears to measure either Type One or Type Two processes seems important. It seems that these, along with the rest of the conclusions of this study, would be essentially disregarded if Q scoring were to continue in the traditional manner.
CHAPTER V

SUMMARY AND CONCLUSIONS

Summary

The subject of Rorschach Q responses - and in fact the whole concept of originality - has received little attention in the psychological literature. Rorschach testers routinely score Qs; the Rorschach Scoring Summary forms include a space and a heading for Q totals; and the literature on projective tests continues to encourage the belief that there is such a thing as a Popularity-Originality dimension in Rorschach data. On the other hand, the texts on projective tests say very little about Q responses; most projective test users appear to have only the most superficial understanding of the possible clinical significance of Q responses; most testing reports do not include any statements about, or conclusions based on, Q totals; and the testing literature contains virtually no research on, or discussion of, the Q scoring category. It is as if no real usefulness has ever been discovered for this scoring category but nevertheless there is a deep-seated reluctance to either investigate its possibilities or to discard it as non-productive. It seems reasonable to believe that the Q scoring category has survived until this time because of a respect for its historical beginnings.

Methodological issues and intriguing clinical questions abound concerning the Q category. It seems that it would be a very fruitful
subject of inquiry. The few studies that have used and reported Q data have tended to report results that disconfirmed common-sense understandings about such data, thus giving cause to research the question further. Instead of additional questioning, though, less questioning has occurred, and for all practical purposes, that is where the matter stands at this time. The Q scoring category may reasonably be viewed as a question in search of an answer.

The present study began with an interest in attempting to specify the nature of the psychological processes that produced Q responses. In the course of administering Rorschachs to many subjects it was observed that there seem to be two distinct ways that an Q score comes into being. One is by associating a truly unique and unusual content to a common blot area. The other is by associating any content at all to a highly uncommon blot area. The criterion for scoring Q is that the response occurred less than one time per 100 responses to the blot area used. This criterion underlies the first of the two ways stated above for generating an Q response. The other way approaches the subject from the opposite direction, by focusing on the statistical infrequency of the blot area chosen rather than on the content of the response. Specifying two separate processes like this seemed to be important, and in an effort to maintain the distinction different names were given to these two processes.

Subjects who respond to infrequently seen blot areas are said to exhibit Type One perceiving or idiomorphic processes. Subjects who produce highly uncommon content to regularly-seen blot areas are said to exhibit Type Two perceiving or idiographic processes. This terminology is unique to this study. The assumption was made that
Type One perceiving could be measured by the Embedded Figures Test (EFT) and that Type Two activity could be measured by the Remote Associates Test (RAT). Then an experimental procedure was devised that was intended to generate maximal numbers of O responses. The procedure involved four major steps. One was to use an objective scoring standard to score all Rorschach responses. Forty Rorschachs were administered (20 male and 20 female subjects) and all responses were scored according to the frequency tables published by Hertz. A second step was to invent a special Testing of the Limits for O responses. This involved interspersing three actual O responses (according to the Hertz tables) with two nonsense responses to twelve blot areas on selected cards and then inquiring as to whether or not the subject would accept any of those responses. The number of actual Os minus the number of nonsense Os was halved and the result was the TOL score. A third step was to designate a new measure or variable, termed TOTAL O, which was the sum of the number of Os given during the standard Rorschach administration plus the TOL score. The net effect of these three steps was to substantially increase the magnitude of the O data that could be used in subsequent statistical operations. This was desired because it was assumed that the number of O responses that would be obtained using the traditional subjective scoring method would be so small and so restricted in range that they would not be amenable to statistical processing. A fourth step that was taken was to maintain separate data headings for the total number of responses to the standard Rorschach administration (#R), for the total number of O responses given during the standard administration (#O), and for the
percentage of Q responses relative to the total number of responses given during the standard administration. These three additional categories were really "in case of..." headings - since this was a pioneering study these data were systematically generated to see if they revealed some information that would prove useful in gaining maximum information about what amounts to a new topic of inquiry.

Three hypotheses were formulated concerning relationships between EFT scores and number of Q responses, between RAT scores and number of Q responses, and between EFT scores and RAT scores. All subjects were tested with the Rorschach (including the new TOL procedure), as well as the EFT and the RAT. The tests were scored by two independent scorers each using the Hertz tables after a period of discussion about similar scorings of pilot data. The data were collected and analyzed, with male and female data kept separately. Correlation matrices for the male, the female, and the combined data were constructed consisting of the rank-order correlations between all the seven data headings of the study. Critical values were determined for all the correlations obtained. A Pearson $r$ was also computed for the Q scorings of the two judges, with a separate $r$ for male and female data.

The data were then discussed in terms of four particular factors. One was the nature of the psychological processes that seemed to underlie the final data of the study. This involved Type One and Type Two perceiving and the inferred mental processes that result in Rorschach Q responses. This related directly to one of the concerns that originally motivated this study, namely an attempt to
determine what clinical significance, if any, the Q scoring category possesses. The second factor pertained to research methodology. There were some unique methodological features of this study that deserved special attention in their own right, regardless of the specific results obtained. Such things as the use of the Hertz tables and the TOL procedure, for instance, required discussion of their effectiveness as research tools. A third factor was the special appropriateness of discussing the concept of originality. This, again, was a major interest in designing this study. The data did not seem to lend themselves to a lengthy discussion of originality, but it is hoped that making this subject the central topic of a research study will encourage further research concerning the concept. A final factor was that this study was in many ways a ground-breaking effort. There was very little related literature and so the data were treated in a way that was intended to cover as much content area as possible. The idea was as much to suggest new ideas, to explicate and to question basic assumptions, and to stimulate interest in the general topic of Q responses as it was to provide specific new data on the subject.

Conclusions

1. There is a statistically significant relationship between performance on the EFT and the production of Rorschach Q responses for female subjects. There is no uniform relationship between EFT scores and any Rorschach Q measures for male subjects. It is concluded that Type One perceiving does foster Q production but the EFT is not an accurate measure of idiomorphic activity. There is no support in the data for any attempt to posit
relationships between production and any of the personality characteristics associated with the concept of psychological differentiation.

2. There is no statistically significant relationship between performance on the RAT and the production of Rorschach Q responses, for either male or female subjects. The RAT evidently does not measure idiographic processes and thus there is no significant uniform relationship between RAT performance and Type Two perceiving. There is also no support for inferring relationships between Q performance and personality characteristics that correlate positively with RAT performance.

3. There is no statistically significant correlation between the EFT and the RAT scores of the female subjects. There is a statistically significant rank-order correlation between the EFT and RAT scores of the male subjects. It is inferred that this finding relates to a general tendency for men to perform more comfortably and more uniformly on relatively well-structured tasks with specific right-wrong answers, while women seem able to respond effectively to a wider range of testing situations. It appears that female subjects generally exhibit a more accepting and less critical, less structure-bound, less literalistic and guarded attitude toward the test than male subjects do.

4. There is statistically significant positive interscorer reliability between two independent scorings of all the Q data, using the Hertz frequency tables as an objective scoring
5. The TOTAL Q measure has a statistically significant uniform relationship with the four other Rorschach measures of this study. This is an unremarkable conclusion because the TOTAL Q heading subsumes the data of three of the other four headings.

6. The five Rorschach measures (#R, #O, %O, TOL, TOTAL Q) tend to intercorrelate at statistically significant levels (nine of the ten correlation coefficients for the data from female subjects are significant, all at the .01 level; and seven of the ten rho values for the male data are significant, with six of the seven significant at the .01 level and the seventh at the .02 level). These results suggest a large degree of overlap between some of these measures. It was concluded that the %O, TOL and TOTAL Q categories are redundant or uninformative and do not need to be maintained as separate headings. This means that the special methodology established for this study in order to generate additional Q responses proved to be unnecessary. It is suggested that #O be replaced by %O as a clinical measure. The #O data of this study indicate that sufficient Q responses to be used in statistical operations could be obtained from a standard Rorschach administration, provided: (1) some objective scoring standard were used in scoring Qs; (2) some special effort was made to encourage male subjects to respond more loosely; and (3) test examiners be trained to outline precisely those areas of the blot that subjects use for each response.

7. The standard Rorschach administration will produce protocols in
which roughly 40% of the total number of responses are Qs, provided that some objective scoring standard is used in scoring all responses, and that the examiner closely records the exact blot area outlined by the subject during the Inquiry phase. This percentage of Q responses can routinely be expected in the protocols of psychologically normal subjects. Whether or not intelligence level is a factor was not tested, although the subjects in this sample were drawn from a population known to be above average in intelligence.

8. The only personality characteristics that were not eliminated from consideration as possible correlates of Q activity were what Barron has termed a preference for complexity and some degree of imbalance in phenomena, and a tendency to entertain ideas and impulses that are commonly taboo. This study did not test for relationships between Q performance and these two personality characteristics, but it did disconfirm a relationship with two distinct clusters of personality features. The two mentioned here are survivors, in a sense, and might serve as profitable beginnings for future research.

9. This study seems to underscore some questions that have remained largely implicit regarding the nature and the usefulness of the Q scoring category. In particular, these data focus on the questions: are Q responses so infrequent that few may be expected in an average record; and do Q responses result from strikingly original mental associations to common stimuli? The answers that this research provides for those questions tend to be
interrelated and to center around the fact that very few responses in these data seemed to result from Type Two activity. Most obtained in this study appeared to be the result of subjects responding to similar, but slightly different, blot areas on the same cards. Since the blot areas are different from those used by other subjects, the responses are likely to be scored even if the content is nearly identical in each instance. This observation has far-reaching implications. It means that a fair number of responses can be anticipated per record since there are a vast number of ways of dividing up any blot area. This study found that an average of approximately 45% of the total number of responses given by any subject, regardless of gender, will be responses. This is a much higher percentage than would be expected on the basis of the traditional assumptions about responses.

The scoring of responses could be limited to responses that only seem to reflect Type Two processing. This would greatly decrease the number of responses obtained per record. This would also bring the results of this study much closer into line with the traditional assumptions about responses. In addition, it would eliminate any concern about altering or even discarding the scoring category. As matters stand now, however, the retention of the category in its present state is open to serious question. This research suggests that the category is not in fact what it has been assumed to be. It is not as limited or as exclusive as previously thought, and it does not seem to be
a measure of psychological processes that are characterized by
the production of unique associations to commonplace stimuli.
Since so little of this latter type of activity occurs, there
seems to be little use for a scoring category to measure it.
If the Q category is to be retained at all, it will probably
have to be understood as a measure of what this study has
termed idiomorphic psychological processes.
REFERENCES


Hertz, M. Frequency Tables for scoring responses to the Rorschach Inkblot Test. (3rd ed.) Cleveland: The Press of Western Reserve University, 1951.


Roe, A. Psychological examinations of eminent biologists. *Journal of Consulting Psychology*, 1949, 13, 225-246. (b)

Roe, A. Psychological tests of research scientists. *Journal of Consulting Psychology*, 1951, 15, 492-495. (a)

Roe, A. A psychological study of eminent biologists. *Psychological Monographs*, 1951, 43, 121-239. (b)

Roe, A. Analysis of group Rorschachs of psychologists and anthropologists. *Journal of Projective Techniques*, 1952, 16, 212-224. (a)


Thorndike, E., & Lorge, I. *A teacher's word book of 30,000 words.* New York: Teacher's College Columbia University, 1944.


APPENDIX A

MATERIALS FOR TESTING OF THE LIMITS FOR RORSCHACH Q RESPONSES

82
Testing the Limits for Q Responses

The information obtained from the procedures that are outlined below will be used for a dissertation project, so please be conscientious about following these instructions. The purpose of the project is to gather data regarding the clinical significance of Rorschach Q responses.

Certain areas of selected cards are considered maximally productive for Q responses. This study will involve testing the limits of those blot areas - which are marked on the Location Chart and are listed on the Work Sheets which accompany these instructions. Testing the limits for these specific areas are designed to elicit Q responses. Every one of the designated areas are to be used - failure to test in every area may invalidate the record.

General Instructions

This phase of testing will be done after the Inquiry and before testing the limits for Klopfer's populars.

You will be given Work Sheets on which to record responses of the subject. Record everything that you and the subject say. Indicate those questions that you ask as examiner, using the symbol: (?).

Please write legibly.

1. Give Card I, in the upright position, to the subject. Ask:

   Q1 When you saw this the first time you said that it reminded you of... (repeat the subject's responses to the card). Now that you look at it again does it remind you of anything else?

   Record verbatim whatever the subject says.
2. Even if the subject has already responded to the area at some previous time, ask:

Q2 When you look at just this area (outline the designated area), what does it remind you of?

Ask this question for each of the 12 areas listed. Record all responses verbatim.

3. Whether the subject has already responded to the area or not, ask:

Q3 Could this be a ...? (Read each of the five responses listed on the Work Sheets. Q3 is to be asked separately for each of the five responses per card. If the subject has already given one or more of these responses before this point, you may eliminate this question.)
Work Sheet: CARD I

$Q_1$: (?) -
Response -

$Q_2$: (?) - Area 6 &/or 7 -
Response -

$Q_3$: (?, pencil -
Response -

(?) - steps -
Response -

(?) - bicycle -
Response -

(?) - glaciers -
Response -
Work Sheet: CARD III

$\Omega_1$: (? -)
Response -

$\Omega_2$: (? - Area W -)
Response -

$\Omega_3$: (?), airplane -
Response

(?) , bird or insect, stepped on, scattered and in pieces -
Response -

(?) , elephant
Response -

(?) , entrance to a park, with trees, snow, bushes, and red flowers -
Response -

(?) , germ: microscopic view of germ, highly magnified -
Response -

$\Omega_2$: (? - area 5)
Response -

$\Omega_3$: (?), building
Response -
(?), pelvis - Response -

(?), pistol - Response -

(?), water wings - Response -

(?), kidneys - Response -

\[ ?_2: \ (\) - area 23) \]
Response -

\[ ?_3: \ (\) , truck making a delivery \]
Response -

(?), lamp - Response -

(?), tropical rain forest - Response -

(?), mushroom - Response -

(?), chalice - Response -
Work Sheet: CARD VIII

Q1: (?) -
Response -

Q2: (?) - area 3+3 -
Response -

Q3: (\?), water buffalo, pulling a plow -
Response -

(\?), corset, old fashioned -
Response -

(\?), cup and saucer -
Response -

(\?), sails, of a boat or a ship -
Response -

(\?), jacket, laced -
Response -

Q2: (? - area 20)
Response

Q3: (\?), automobile -
Response -
(?) filament of an electric light, with the wirey part -
Response -

(?) attache case -
Response -

(?) Chinese pagoda -
Response -

(?) totem pole -
Response -
Work Sheet: CARD IX

\[ \mathcal{Q}_1: \ (\ ?) \ - \\
\text{Response} - \]

\[ \mathcal{Q}_2: \ (\ ? - \text{area W}) \\
\text{Response} - \]

\[ \mathcal{Q}_3: \ (\ ?), \text{drug capsule} - \\
\text{Response} - \\
(\ ?), \text{monster with big claws, climbing up from clouds} - \\
\text{Response} - \\
(\ ?), \text{microphone} - \\
\text{Response} - \\
(\ ?), \text{persons in costume, with bright colored clothes, facing each other} - \\
\text{Response} - \\
(\ ?), \text{witches dancing amid clouds} - \\
\text{Response} - \]
Q₁: (?) -
Response -

Q₂: (?) - area W)
Response -

Q₃: (?), a platter of bacon and eggs -
Response -

(?, elves in fairyland, in color -
Response -

(?) , a Zodiac chart -
Response -

(?), an aquarium -
Response -

(?) , fireworks -
Response -

Q₄: (area 3+3) -
Response -

Q₅: (?) , a cavalry charge -
Response -
(?) a person's moustache -
Response -

(?) a Christmas tree -
Response -

(?) a worm on a hook -
Response -

(?) a wishbone -
Response -

\[ Q_2: (? - area 6) - \]
Response -

\[ Q_3: (?) , weather balloon - \]
Response -

(?) witch on a broomstick -
Response -

(?) basketball net -
Response -

(?) snowflake -
Response -

(?) seaweed -
Response -
Q2: (? - area 7 &/or 52) -
Response -

Q3: (?), an anvil -
Response -

(?) fried egg, "sunny side up", with yolk in middle -
Response -

(?) ship's anchor -
Response -

(?) biology slide, showing cell structure -
Response -

(?) canary -
Response -

Q2: (? - area 14)
Response -

Q3: (?), a key chain -
Response -

(?) maple seed pod -
Response -

(?) telephone -
Response -
(?), door knocker -
Response -

(?), pawnbroker's sign -
Response -
The Dissertation submitted by Joseph Frederico has been read and approved by members of the Department of Psychology.

The final copies have been examined by the director of the Dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the Dissertation is now given final approval with reference to content and form.

The Dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

November 14, 1973
Date

Frank Wuln
Signature of Advisor