INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in “sectioning” the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.
SELING, MARK J.

SYNCRISIS: INVESTIGATIONS OF A NEW ASSESSMENT PROCEDURE

Iowa State University

University Microfilms
International

300 N. Zeeb Road, Ann Arbor, MI 48106

18 Bedford Row, London WC1R 4EJ, England
Syncrisis: Investigations of a new assessment procedure

by

Mark J. Seling

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

Major: Psychology

Approved:

Signature was redacted for privacy.
In Charge of Major Work

Signature was redacted for privacy.
For the Major Department

Signature was redacted for privacy.
For the Graduate College

Iowa State University
Ames, Iowa

1980
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>ASSESSMENT AND PSYCHOMETRICS</td>
<td>2</td>
</tr>
<tr>
<td>PSYCHOMETRIC ANALYSIS</td>
<td>3</td>
</tr>
<tr>
<td>The Idea of Similarity</td>
<td>9</td>
</tr>
<tr>
<td>Psychological Diagnosis</td>
<td>16</td>
</tr>
<tr>
<td>SYNCRISIS</td>
<td>21</td>
</tr>
<tr>
<td>Syncrisis vs Scale Scoring</td>
<td>22</td>
</tr>
<tr>
<td>Syncritic vs Scale Assessment</td>
<td>23</td>
</tr>
<tr>
<td>Other Uses and Limitations of Syncrisis</td>
<td>24</td>
</tr>
<tr>
<td>Studies of Syncrisis</td>
<td>27</td>
</tr>
<tr>
<td>EXPERIMENT 1: RELIABILITY OF SYNCRISIS</td>
<td>28</td>
</tr>
<tr>
<td>Method</td>
<td>30</td>
</tr>
<tr>
<td>Sample</td>
<td>30</td>
</tr>
<tr>
<td>Procedure</td>
<td>30</td>
</tr>
<tr>
<td>KOIS</td>
<td>31</td>
</tr>
<tr>
<td>JPI</td>
<td>33</td>
</tr>
<tr>
<td>Subjects and criterion people</td>
<td>34</td>
</tr>
<tr>
<td>Analyses</td>
<td>35</td>
</tr>
<tr>
<td>Open-pool analysis</td>
<td>35</td>
</tr>
<tr>
<td>Fixed-pool analysis</td>
<td>39</td>
</tr>
<tr>
<td>Results</td>
<td>40</td>
</tr>
<tr>
<td>Discussion</td>
<td>46</td>
</tr>
<tr>
<td>EXPERIMENT 2: CLIENT-THERAPIST MATCHING</td>
<td>51</td>
</tr>
<tr>
<td>Matching Research</td>
<td>54</td>
</tr>
<tr>
<td>Method</td>
<td>63</td>
</tr>
</tbody>
</table>
**EXPERIMENT 3: OUTCOME OF ACTUAL COUNSELING AS A FUNCTION OF CLIENT–THERAPIST SIMILARITY**

### Method

- Sample
- Procedure
- Predictive methods and outcome variables

### Results

### Discussion

### CONCLUSIONS

- Criticism 1: Syncrisis Produces Impoverished Data
- Criticism 2: Definition of "Content Relevance"
- Criticism 3: Matching People Implies that "It Takes One to Know One"

### REFERENCES

### ACKNOWLEDGMENTS

### APPENDIX A

### APPENDIX B
INTRODUCTION

syn·cri·sis (singˈkri sis, sinˈ-), n.
Rhet. Obs. 1. a critical comparison.
2. the act of a critical comparison.
[<GK syn- with, together + kri( nein) (to) decide, separate + sis]

This dissertation involves theoretical discussion and empirical study of a new assessment procedure termed "syncrisis". Briefly, the method is in some ways analogous to a matching procedure, and it is asserted that syncrisis represents a potentially fruitful development in assessment technology. Elements of the new method were first suggested by Frederick Kuder (1977a,b), so there is little literature on the topic per se to review. Rather, the task of this thesis is to articulate the theoretical reasons for using syncrisis, and then investigate the properties of the procedure. In order to describe adequately the concept of syncrisis, it is first necessary to discuss a number of related background issues. Following the introductory chapters, studies of the psychometric properties of syncrisis are described separately.
Peoples' interest in assessing each other seems interminable. Attempts to assess people, or size them up, most certainly predate the emergence of psychology as a discipline, the appearance of which is usually fixed in the 1870's (e.g., Boring, 1949). In fact, entire systems, well-articulated if naïve, for performing assessments have been around since antiquity (cf., McReynolds, 1975). Astrology and phrenology, for example, represent two prescientific, thoroughly organized systems for providing assessments. The two schemes even have in common with applied psychology the general aim of gaining insights into people through systematic means. The purposes for seeking such an understanding may range from a need to merely satisfy curiosity about oneself, to needs for information to assist the processes of career planning, psychotherapy, and education. At about the time of the identification of psychology as a discipline, the related fields of statistics and psychological measurement were launched with the works and talents of scientists like Galton, Pearson, and Thorndike (Watson, 1971). With these developments, psychological assessment with objectively scored inventories departed from the prescientific tradition. Early assessment devices, "structured" inventories as they came to be called, were little more than collections of common sense questions concerning personal disposition (e.g.,
Bernreuter, 1931) or vocational interests (e.g., Thorndike, 1912). Subsequently, psychologists like Strong (1943), and Hathaway and McKinley (1942) produced inventories which placed much emphasis on empirical verification of assessment by inventory (Hathaway, 1965; Campbell, 1968), for example, results of assessments were tested for validity. The branch of psychology known as psychometrics resulted from these influences, and it is with issues of both assessment and psychometrics that we are presently concerned.

Assessment, sizing people up, and psychometrics, measurement of psychological attributes are intimately intertwined (e.g., Helmstadter, 1964; Cronbach & Gleser, 1965; Edwards, 1970; Wiggins, 1973; Nunnally, 1967, 1978; Goldman, 1971). While it is possible to identify numerous similarities between the two, it is important to note the important distinctions between them. Assessment is a judgment, evaluation, or inference (e.g., "psychopath", or "genius"); it is a comment, or a summary picture (McReynolds, 1975). Psychometrics is concerned with measurement, or assignment of numbers according to rules (Nunnally, 1978); it is a means for providing quantitative figures to single features a person has (e.g., degree of hostility, or intelligence quotient), usually through testing. Goldman (1971) likens the relationship between a test (psychometrics) and an interpretation (assessment) as being connected by "bridges". His analogy has been
adapted to fit the present distinction between psychometrics and assessment to derive Figure 1.

The figure depicts the relationship as one of interdependence. Thus, an assessment may be based upon psychometric tests. There remain, however, a few further points to be made.

It was stated above that psychometrics is concerned with measurement. As such, the term may be used in a variety of ways, e.g., "psychometric device" (a test, or inventory), "psychometrics" (study of quantitative methods), or even "psychometric properties" (qualities of a test; reliability, validity). Therefore, the relationship between assessment on the one hand, and psychometrics, on the other, is not always as it is described by Figure 1. Specifically, it is clearly possible to study psychometrics qua psychometrics, at least without the goal of assessing a specific person.

![Figure 1. Relationship between psychometrics and assessment (after Goldman, 1971)](image-url)
This thesis, for example, concerns the psychometric properties of two inventories when they are used or scored in different ways. The question addressed is not one of whether one or the other inventory is better, nor what kinds of clinical assessments based upon the devices are justified. Rather, the issue is what are the psychometric properties of a new method for analyzing peoples' responses to inventory questions compared to standard methods. Each method has its own implications for the problem of assessment.
PSYCHOMETRIC ANALYSIS

The typical test, regardless of how it was developed usually allows for a score (or, scores) to be derived to permit the individual to be somehow compared with other persons, the criterion group. The technology for developing personality and interest inventories allows a distinction to be made between homogeneous and empirical keying strategies (these may be combined, e.g., Jackson, 1976), but nonetheless, a person's score on a test nearly always involves a comparison to other individuals. To illustrate, consider a test developed to measure anxiety (homogeneous key) or to identify anxious people (empirical keying).

The former strategy begins with the identification of items or questions which seem rationally to be related to anxiety (c.f., Jackson, 1971). A large number of such items are then administered to a sample of people and then, in the simplest case, items are retained for the final scale if they correlate highly with one another and with total test score. Selected items are standardized on another sample—a mean score is determined, degrees of deviation are specified, etc. When the final test is given to someone, an "anxiety" score is derived from comparing his or her responses to the average responses of the persons used in the derivation samples. The earned score is thus at, above, or below the mean of such a group. This permits the conclusions that the subject
is more or less anxious than average, or that one of two subjects who have taken the test is more or less anxious than the other.

The empirical keying approach is somewhat different in item selection but the final results are comparable to that described above. Items are retained for final use on such a test only if they have a high correlation with a criterion. Thus development of such a test begins with identification of a criterion group, anxious persons to continue the example, and a control group. The two groups are then given the potentially useful items (which may or may not appear to measure anxiety) and items are studied to see which of them have differential endorsement rates in the two groups. Items which are answered in proportionally similar ways by the criterion and control group are not used on the eventual scale. Scores on the final scale are then standardized, as described above, and a subject's taking the final product permits conclusions similar to those listed for a homogenous key.

There was for a time in the field of psychometrics a widespread debate over the relative merits of each keying strategy (cf., Meehl, 1945; Jackson, 1971). Subsequent research (Hase & Goldberg, 1967; Burisch, 1978) has shown that both strategies have about equal validity. However, consensual opinion is now that the most useful approach is to use both technologies so that the resultant scale has both con-
Struct validity (measures the construct—e.g., anxiety—that it purports to) and predictive validity (accurate identification of, say, anxious persons). See Nunnally (1967, 1978) for a discussion of this.

Assessment of individuals based on such standardized instruments is presently in widespread use. There are literally thousands of these kinds of instruments available (Buros, 1978). The purpose of describing the empirical and homogeneous keying strategies in trait measurement has been to emphasize the way in which psychologists have adopted the measurement model described in the last chapter. That is, testing involves comparison of an individual's score to a standard, which is formed on the basis of average item endorsement tendencies of a large, test derivation group. In order to compare two individuals on some attribute (or collection of attributes) it is first necessary to perform the intermediate step of comparing each of them to the standard. It is my thesis that there are times for which more direct comparison of individuals would be useful. One might, for example, directly compare peoples' answers to test items and determine a similarity score based on agreement and disagreement in item level response patterns. This is the process of psychometric syncrisis. The result of such a procedure could be an assessment of a person based on a series of individual comparisons with other persons concluding that the testee is
most nearly like one or the other. In fact, the comparison (or, criterion) group could all be rank ordered for degree of similarity to the testee, and from a knowledge of what the criterion people were like, an assessment of the testee could be developed. This is syncritic assessment.

Both syncrisis and standard keying strategies involve the process of making comparisons. In syncritic analysis, the comparisons are between individuals, while in trait assessment the comparisons are between a testee and a standard arrived at by scaling the responses of a criterion group. In either case, the result of a comparison is a judgment about the relative similarity of the compared entities. The concept of similarity, then, is fundamental in psychological assessment. Because syncritic analysis is being offered as an alternative to standard assessment methods, it is my obligation to spell out the relationship between similarity on the one hand, and syncrisis on the other. The next section is intended to meet this end.

The Idea of Similarity

"Similarity" may be used in a variety of ways: logical, classification, and feature similarity are among the possibilities (Gregson, 1975). The previous section of this thesis identified the process of assessment as one involving comparison and judgments of similarity between people, whether con-
sidered in pairwise fashion between two people, or with reference to a standard, averaged group response tendency. While there are other uses of the concept of similarity in psychology, for instance, multidimensional scaling of objects to determine what it is about them that guides subjects to classify them in certain categories (Nunnally, 1978; Krumhansel, 1978; Tversky, 1977), the present concern is on similarity between people as may be inferred from their responses to psychological inventories.

Such responses to psychological inventories may be analyzed in a variety of ways. Because psychological similarity is inferred from response similarity, the inference(s) to be made will depend on the kind of psychometric analysis performed. Various means of conducting such an analysis, and their consequences, are described below. A related means of determining psychological similarity, that of assigning people to taxonomic groups, is lightly touched upon.

The context of this discussion of similarity concerns analysis of responses to items on psychological inventories. It follows that there are two possible levels for such analyses, item level and profile level. The first of these involves direct use of answers to questions on the test in question, without reference to scores on particular scales for which an inventory may be scored. The second method, profile level analysis, ignores responses to individual items
and proceeds to determine degree of similarity on the basis of amount and pattern of differences between peoples' scale scores. Ordinarily, these approaches are not combined since so doing would add redundancy to the analysis, and possibly undesirable consequences.

Profile similarity analysis begins with the recognition that profiles can show three distinct characteristics (Cronbach & Gleser, 1953): scatter, shape, and elevation. Figure 2 shows the profiles of subjects tested with a five scale inventory. Elevation, or level, refers to the average score of the person over the mean score of the scales on the profile.

Figure 2. Characteristics of profiles (see text)
subject A has the greatest elevation. Scatter, or dispersion, is represented by the variability of a single person's scores over all scales, subject B shows the least dispersion. Shape of the profile is the pattern of elevation across the scales of the inventory, taking the order of these scales into consideration (i.e., leaving order the same, or constant). Subjects A and C in Figure 2 have most nearly the same shape.

At the profile level, the problem for assessing similarity is to take these features into account when calculating a similarity index. It is not easy to account for all three features with one index (Alenderfer, 1977, Note 1; Sneath & Snokal, 1973; Greenhouse & Geisser, 1959; Harris, 1955; Cattell, 1949; Skinner, 1978). There is cohesive and extensive literature on this subject unto itself, but the discussion below is limited to two of the more widely used indexes without pretense of reviewing the literature on profile similarity indexes.

The first method for determining profile similarity is linear distance, or Cronbach and Gleser's $D^2$ (1953). The expression for the distance between two subjects, A and B, may be written as:

$$D_{AB}^2 = \sum_{k=1}^{N} (A_k - B_k)^2$$

This determines the sum of squared differences between two subjects on $N$ scales. Referring to Figure 2, A and C have the greatest distance between them. Thus, $D^2$ attends to all three
profile characteristics. In cases where it is desired to avoid reflecting a major contribution due to elevation, profiles may themselves be standardized.

A second means for assessing profile similarity is to correlate subjects' scale scores. Correlation coefficients attend mostly to shape. Subjects A and C would be determined to be most nearly similar with a correlation similarity index. The primary problem with the correlation coefficient is that it can be high even if two profiles do not have the same shape as long as they are linearly related (Fliess & Zubin, 1969). With the use of either $D^2$ or correlation, as with any summary procedure at the profile level, there will be a loss of some information.

One way to prevent this information loss is to examine item responses directly, although item analysis can present problems of its own. Primarily, the problem is that individual responses to single questions are not as reliable as scale scores based on many items (Nunnally, 1978). It can be argued on the other hand (Borgen & Scott, Note 3; Kuder, 1977a), that while individual items may not be highly reliable the vector representing the responses to all items on an inventory is itself quite reliable.

There is a second potential pitfall for assessing similarity based on responses to items. (An attempt to avoid this problem led to the selection of the two inventories
described in the Methods sections for use in the present studies.) The problem concerns the content validities of the test items. Content validity refers to how well the items on a test represent or reflect the domain of interest (Betz & Weiss, 1976; American Psychological Association Standards for Educational and Psychological Tests, 1974). The extent to which this is accomplished can be determined by expert judgment or internal consistency analysis. Meeting this aim always begins with an awareness of the issue, and both Kuder (1974) and Jackson (1976), the authors of the inventories which were used in this dissertation, devoted extensive effort to widely sampling the domains measured by their inventories.

The importance of this arises for the following theoretical reason. If two persons are said to be similar, or alternatively more similar to each other than to a third person, the crucial issue is in what way are they similar? In order to understand the results of a matching of individuals (i.e., the process of determining how similar they are) a knowledge of the test content is necessary. At the extreme, if items were randomly selected and two peoples' responses were compared, no conclusion could be generalized about their similarity. All that could be said is that they did or did not respond in comparable ways. Because similarity is an inference, the basis for such generalizations—that is, the test content—must be clearly defined.
If the content of an inventory is clearly defined then the possibilities for inference are also clear. For example, a test measuring a single content area, such as anxiety, can lead to conclusions that are somewhat different than those identified previously. Specifically, the conclusions permitted are that two individuals are (are not) similar in amount of anxiety, and they are, in turn, similar or not similar to other persons. Also possible is a rank ordering of several other people compared to the single one. If it is necessary to determine how much anxiety a person has, then a comparison to the scale standardization norms would have to be made. Alternatively, rather than referring to norms, one could refer to other data available on the criterion subjects to get an understanding of the testee. This alternative was suggested by Kuder (1977a, b). Consider a subject who has taken an interest inventory. The normative approach allows the conclusion that his or her interests are most nearly like, say, lawyers, and accountants. The item level similarity approach allows for the possibility of concluding that the subject's responses were most nearly like some number of individuals who are engaged in such and such occupations. Removed is the intermediate step of comparing peoples' responses to norms. The implications of removing this step are described in the next chapter, which also discusses specific details about how syncrisis is performed. Before turning to the how and what
Psychological Diagnosis

Wade and Baker (1977) recently published a survey of clinical psychologists occupational activities. A major finding of their survey was that fully one-third of clinicians' service delivery time was spent on assessment. This indicates that assessment activity is considered quite important in applied settings. It is one task of basic research to provide applied psychologists with effective assessment procedures. The need for such technology has been stated previously: The insights or understanding of the assessee are desired to assist the process of education and psychological treatment. This remains the ideal, but there have been a number of problems with applying these ideals to practical situations.

One problem with traditional assessment has been that it has never been shown that diagnosis is in fact related to type of treatment given (Sundberg, Snowden, & Reynolds, 1978). This is often said to be due to problems inherent in the American Psychiatric Association Diagnostic and Statistical Manual (DSM II, 1968) system of classification. Among the faults of DSM II are the use of broad, poorly defined categories, lack of specific inclusion rules, and lack of a theory to relate
diagnosis to treatment (c.f., Millon, 1969). Although DSM II
will be soon superseded by DSM III (Spitzer et al., 1978)
there have been prepublication criticisms of the new system
(Schact & Nathan, 1977).

In addition to problems particular to specific systems,
such as DSM II, there are also consequences of diagnosis which
may be undesirable. Chief among these are the perjorative
results arising from labeling a person "schizophrenic", or
"retarded" (cf., Rosenhan, 1973; Journal of Abnormal Psy­
chology, vol. 84, 1975, entire issue). The intention of brief,
descriptive labels is to provide efficient communication
(Blashfield & Draguns, 1976; Woodruff, Goodwin, & Guze, 1974).
However, such intention has been thwarted by the fact that
people in society at large tend to have negative opinions of
people with psychological impairments, thus being labeled can
lead to prejudicial treatment. Unfortunately, some psy­
chologists are also influenced by labels. Langer and
Abelson (1974) have embarrassed the profession by showing
that psychoanalysts' judgments of the same person on video­
tape were influenced by whether they were told the interviewee
was a job applicant or a mental patient.

Although many people have assumed that diagnostic
activity would be abandoned in the face of such disasters,
there are reasons why it has not declined. One of these rea­
sons must be that people still adhere to the ideal, seemingly
so successful in the practice of medicine, that accurate diagnosis should facilitate appropriate psychological treatment. Some psychologists, namely the behaviorists, have in fact given up on traditional psychological assessment. In its place they have substituted "behavioral analysis" (e.g., Ciminero, Calhoun, & Adams, 1977; Cone & Hawkins, 1977; Hersen & Bellak, 1976). Although the method is more direct (i.e., observation of a person's actual behavior) the goal remains essentially the same, viz, identify the problem and implicate a treatment for it. The behavioral analysis solution is presently being widely researched. Other solutions may be developed in the future as investigators strive to improve psychological treatment technology.

Structured inventories have been widely used to identify people's pathologies and classify them accordingly. The Minnesota Multiphasic Personality Inventory (Hathaway & McKinley, 1942), one of the most frequently used inventories in published research (Brown & McQuire, 1976), is a prime example. Not only are there rules for analyzing profiles to determine whether a respondent fits one of the various DSM II categories, the inventory is the basis for a widely used classification scheme of its own (Dahlstrom, Welsh & Dahlstrom, 1975). It is to the credit of the Minnesota group that their inventory has been found so useful in practical and research applications. There is no doubt that work of
this variety will continue, and one hopes that eventually valid, useful taxonomic schemes of psychopathology will result.

Traditional assessment, with its emphasis on assigning personality constellations to categories, seeks to gain understanding of an individual through knowledge of the group to which he or she belongs. Traditional testing has aimed to facilitate the realization of this goal. However, a big problem for test makers has been the identification of adequate criterion groups upon which to develop scales to be used for predicting group membership. One problem, for instance, is the heterogeneity of response patterns in many proposed criterion groups. When the responses of members of a criterion group are widely different to a set of items an accurate key cannot be developed on the group (Kuder, 1977a). A further critical problem is that there often are not adequate numbers of people in the total population fitting a criterion definition. In 1979, for example, it would not be possible to develop an Astronaut or Female Airline Pilot key for an interest inventory. There are not enough people in these occupations to study the response patterns of a large (typically, 300 people are sampled) subsample in order to permit an accurate key to be made.

This relationship between assessment, with the aim of understanding the testee, and testing, with its requirement
that averaged group responses be the standard to which persons are compared, has influenced the uses of tests. Reference to groups' response patterns seems everywhere to be required. Syncrisis is proposed as a potential means of avoiding some of the pitfalls of traditional assessment (no label, or type, results from it), as a means for perhaps more directly meeting the aims of traditional testing and assessment (e.g., the "career matching" concept of Kuder, 1977a,b), and as a method for meeting ends not easily met through the usual test scoring procedures (see Client-Therapist Matching chapter).
The concept of syncrisis began with Kuder's (1977a) publication of his "career matching" concept. He recognized that the degree of homogeneity of an occupational groups' interests is related to the number of cases needed to develop an empirical key for that group. High variability groups require large sample sizes to establish a stable key for that group. On the other hand, low variability groups, often specialties or subarea of practice, require fewer cases to develop a key. It appears for the low variability group that interest patterns are more congruent among members and that, accordingly, a subject's scale score on a key made from such a group would be more meaningful as an index of similarity than a score comparing him or her to a diverse group. From these premises, it follows that the most meaningful score is one which compares counselees directly to individual people engaged in various occupations.

As it turns out the notion of person-to-person matching of careers is an instance of the more general notion, developed here, of syncrisis. The purpose of the present chapter is to tie together the general concepts of syncrisis, which have been somewhat loosely spelled out in previous chapters.
Syncrisis vs Scale Scoring

Psychometrically, syncrisis refers to comparative scoring of psychological inventories for two individuals at the item level. Thus, the resultant index expresses the degree of congruence of two persons' response patterns across all of the items of an inventory. Since this involves no intermediate step of first converting responses to a scale score, it cannot be said that calculation of $D^2$ or correlation indices of agreement at the profile level (see earlier section on Psychometric Analysis) constitutes syncrisis. Only item level analysis is termed syncrisis.

The purpose of item level scoring is to maximize the amount of information available in the final syncritic score. To understand how item level analysis maximizes the use of information in subjects' responses to questions consider the following simple, hypothetical situation. A test, P, has 10 true-false items. For the first five, an answer of "true" is in the scale keyed direction, while the obverse is true for the last five. Assume further that five keyed responses constitutes an average score. If a subject, A, answers the questions with a pattern of $(T,T,F,P,T,\ldots/F,T,T,F,T)$, he earns a score of five points which is converted to a score of average. A second person, B, responding to the same inventory might answer $(P,T,F,P,\ldots/F,F,T,F,T)$, also resulting in a score of average. At the profile level, the $D^2 = 0$. At the item level,
by contrast, there is only 80% agreement. As a result of the syncritic analysis the differences between A and B are highlighted. Consider a third subject, C, who responds (T,T,F,F, T,//F,T,T,F,T), a pattern 100% congruent with subject A. With respect to subject A, subject C has responded more similarly to A than has B. A difference has been detected among three subjects where none was found at the scale score level. Removing reference to the group allowed detection of these differences.

Syncritic vs Scale Assessment

Based on the assumption that the content domain of some dimension, X, has been adequately sampled, it could be said, to continue the example above, that A and C were most nearly alike with respect to X. (The importance of the adequacy of domain sampling was previously discussed.) If X is the domain of occupationally related interests, and if A were a client seeking career guidance, then a case history of C's vocational activities could be given to A for study as a potential career of interest. This differs from the standard situation wherein the usual result is a high score on one or more occupational scales. The client may then seek out further information about those occupations to which he or she was most similar. On the other hand, when the scoring has been done syncritically, the scores represent similarities to specific individuals. In
this case, it seems desirable to give the client information about the similar people as well as their occupation. Naturally, the context of this example is quite limited in that there are only two criterion people. Assessment via syncrisis takes place within the context of having available a large number of criterion people with completed case histories on each of them.

What if domain X is a personality attribute? Then the rank ordering of criterion people to the client would result in identification of those people to whom the client is most (and least) similar. The counseling psychologist in practice would have to be knowledgeable about these criterion people in order to make any use of these results. One way to accomplish this would be to administer to every client seen at least one inventory in common. Then each new client could be syncritically compared to those previously seen and inspection of former clients records might provide useful information about how therapy could best proceed. Like any assessment, conclusions regarding treatment are subject to revision in light of future developments between therapist and client.

Other Uses and Limitations of Syncrisis

As a scoring procedure, syncrisis seems to be applicable to those instances where individual comparisons are required, rather than diagnostic types. This feature is an asset and a
limitation. By way of assets, the procedure allows use to be made of structured inventories in other than traditional fashion or for applications where standard scoring has not proven useful. One such possibility is that of client-therapist matching, described below. Another might be in the formation of new typologies, as by cluster analysis at the item level (Borgen & Scott, Note 3).

Less obvious, in the context of a discussion of assessment and psychometrics, is the possibility that syncrisis may be useful as a basic methodological tool for general research purposes. Educational psychologists, for example, occasionally perform "aptitude by treatment interaction" research (cf., Cronbach, 1975). Cronbach reports in his 1975 article the results of many studies performed by him and his colleagues. In one study, introductory psychology students and instructors filled out a behaviorist-humanist (BH) scale at the start of the course. A total BH score was then computed for the students and instructors. At the end of the term it was found that the closer the students' total BH scores were to the score earned by the course instructor, the higher the students' grades were for the term. As an alternative measure of student-instructor similarity, the students in such an investigation could be syncritically compared to the instructor (i.e., by using the instructor's answer sheet as the scoring key). The similarity data could then be related to the outcome
(course grade) data as before. Since the independent variable in this sort of study is the degree of similarity between people, it seems that syncrisis might be useful in that it more directly gets at interpersonal similarity by doing away with the intermediate steps of referring to scale scores and inferring similarity from the total score level.

There are several identifiable limitations to syncrisis as an assessment technique. If a diagnostic label is required, for instance, as when a clinician fills out third party payment (insurance) forms, the method is not useful. It is also not useful on projective instruments (e.g., the Rorschach Inkblots) because responses to such inventories are not standardized. There are too many possible variations to allow accurate determination of similarity. Finally, a major limitation might be its lack of utility in ability (intelligence, achievement) assessment. Ability assessment seems to require comparison of the testee to a standard, whether the standard is a norm or a criterion. Since syncrisis does not result in comparison to standards, application of the technique in this area may not be useful. Syncrisis might, however, be of use in determining differences between individuals who have earned similar total, or composite scores.
Studies of Syncrisis

The empirical components of this dissertation are tests of the reliability and validity of the syncritic method. The reliability study is reported in the next chapter. Following that, a second experiment tests the validity of syncrisis for making predictions about peoples' preferences for counselors. The second experiment requires a separate chapter describing the rationale, method, and results of using syncrisis in a client-therapist matching context. After that chapter, a second test of the validity of syncrisis for client-therapist matching is described on the basis of analysis of actual counseling outcome data. These later data were given to the author by Professor Donald G. Zytowski. While analyses of these data was not a part of the original dissertation proposal, the author found some useful insights about the application of syncrisis to client-therapist matching during the course of studying Dr. Zytowski's data. Therefore, a brief, but explicative, chapter reporting some of the results of my analysis of his data follows the chapter on people's preferences for counselors.
EXPERIMENT 1: RELIABILITY OF SYNCRISIS

Any psychometric technique, syncrisis most certainly included, must meet acceptable standards for quality (American Psychological Association Standards for Educational Psychological Tests, 1974). One such standard is reliability, of which one form is dubbed test-retest. This involves administering the same test twice and correlating the results. Experiment 1 obtains such data for two inventories when they were scored syncritically. The two inventories are the Kuder Occupational Interest Inventory (KOIS; Kuder, 1974) and the Jackson Personality Inventory (JPI; Jackson, 1976).

Determination of whether a scoring method shows acceptable levels of reliability involves a judgmental process. Therefore, the reliability of syncritic analysis will be compared to two other sources of reliability data. The first source is the inventory manuals themselves. Each testmaker has provided information about the scale-level reliabilities for his inventory. Reference to these previously published data will permit comparison of syncritic reliability to the reported scale-level reliabilities for the two instruments. Of interest is the question of whether syncrisis can show levels of reliability similar to the high levels shown by the JPI and KOIS scales.

A second source of reliability data was collected for the
purpose of assessing the relative reliability of syncrisis. The technique employed was to correlate subjects' entire profiles from the first and second test administrations (see Method). This procedure provides an index of the overall reliability of a psychological inventory. Somewhat unfortunately, this was only done for the JPI. The reason similar data were not examined for the KOIS was that the keys for each of its over 160 scales are kept a "trade secret" by the organization through which scoring is ordinarily obtained. At the time this dissertation was proposed, arrangements had been made to provide the author with computer scored KOIS data. This arrangement was not carried out. However, the lack of these data is not crucial to the present investigation, where the central issue is the reliability of syncrisis. With the available, previously published scale-level reliabilities adequate judgments of the performance of syncrisis can be made. The primary reason for calculating the profile-level reliabilities is to examine the stability of the answers of the subjects employed in the present experiment, relative to the samples used by the test makers as reported in their manuals. Since this comparison is not central for the determination of the reliability of syncrisis, it was believed by the author that profile-level stability need not be calculated for the KOIS.
Method

Sample
Subjects were 192 male and female undergraduates who volunteered to participate in exchange for course credit in various psychology courses at Iowa State University. The median age for both males and females was 19. There were 16 dropouts, and nine were rejected for high infrequency scores on the Jackson Personality Inventory. Thus, the data reported below are based on 167 subjects for the JPI (74 males, 93 females), and 176 for the KOIS (79 males, 97 females).

Procedure
Subjects were informed that they were to take the Kuder Occupational Interest Survey (KOIS) and the Jackson Personality Inventory (JPI) two times each, separated by a three-week interval. The KOIS and JPI were given in a single session both times to four groups of about 50 subjects. Order of presentation of the two inventories was counterbalanced. People participating in this study were told that the experimenter was interested in analyzing the statistical properties of the instruments. They were told further that if anyone was interested in interpretations of their responses to the inventories these were available a week after the study was finished.
KOIS

The KOIS (Kuder, 1974) is a 100 item empirically keyed inventory which is primarily used to assess occupational interest patterns of test takers. At present, the inventory is scored on 157 scales (Female: 32 occupational, 19 college major; Male: 77 occupational, 29 college major). All 100 items are used in calculating scores on each scale of the KOIS when it is scored in conventional fashion, which is done at Science Research Associates, Chicago. The actual keys themselves are a trade secret.

The scores reported on the inventory are, in effect, point biserial correlations between the test taker’s responses to the 100 items and the averaged responses of a large number of persons selected as representative of the various occupations and college majors. Thus, the score for a given scale is a figure between 0-1.0, with negative figures occurring only when random, or misguided responding has happened. Figure 3 illustrates the format for a KOIS item.

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>PREFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Most</td>
</tr>
<tr>
<td>Activity A</td>
<td>1</td>
</tr>
<tr>
<td>Activity B</td>
<td>2</td>
</tr>
<tr>
<td>Activity C</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 3. The format for KOIS items
Since takers are instructed to respond with one "most" and one "least" preference, there are six possible answers to each item (vis: 1-5, 1-6, 2-4, 2-6, 3-4 and 3-5). This is important to note since it necessitates the use of a weighting scheme when performing syncritic analysis. For example, it is reasonable to assume that of three persons who answer an item with a) 1-2, b) 1-3, and c) 2-1, the first two peoples' answers are more congruent than either the first or second is compared to the third. The weighting scheme employed in this study was developed by Hornaday and Kuder (Note 2). It is: 9 (same "most" and same "least" response), 7 (same "most" response), 6 (same "least" response), 2 (for the two cases when "most" or "least" is one step out of line with the subjects', e.g., if S gives 1, 5 = A, C, B ranking, and second subject gives 2, 6 = B, A, C; or 3, 4 = C, B, A ranking), 0 (no communcality).

The figure representing the total amount of inventory response agreement between two people, hereafter called a syncritic score, could theoretically range from 0-900 since there are 100 items. In practice, syncritic scores for the KOIS run from about 300-650. There is no standard error for this index, hence "significance" cannot be tested. The important result from a syncritical analysis is the rank order of syncritic scores for "criterion" to "subject" comparisons. Figure 4 shows the results of syncritic analysis for a subject
### Figure 4. Syncritic analysis with KOIS items (see text)

<table>
<thead>
<tr>
<th>Subject Response</th>
<th>Criterion People Response: (Weight)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jones</td>
</tr>
<tr>
<td>1</td>
<td>1,5</td>
</tr>
<tr>
<td>2</td>
<td>2,4</td>
</tr>
<tr>
<td>3</td>
<td>2,6</td>
</tr>
<tr>
<td>4</td>
<td>3,4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Syncritic Scores:</th>
<th>(20)</th>
<th>(27)</th>
<th>(10)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Rankings:</th>
<th>Person</th>
<th>Syncritic Score</th>
<th>(total possible)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jones</td>
<td>36</td>
<td>(total possible)</td>
</tr>
<tr>
<td></td>
<td>Wundt</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wolpe</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Freud</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

compared to three criterion people on a hypothetical (4-item) KOIS.

**JPI**

The JPI (Jackson, 1976, 1978) is a 320 item, homogenously keyed inventory which yields scores on each of 15 personality dimensions. Each scale has 20 items, and an infrequency scale is based on the remaining 20 items. The items are all true-
false, with 10 on each scale scored true. Scores for each scale are given in the form of T-scores ($M = 50$, $SD = 10$). The impressive procedures employed for item selection and the large validation samples ($n = 4000$) were reasons for selecting this instrument; the minimum required for use in this study was a personality inventory with true-false (dichotomous) items.

Syncritic analysis of JPI response sheets consisted of calculating proportion of agreement for each subject compared with all "criterion" peoples' responses. Ranking of criterion people was accomplished through ordinal arrangement of the proportion agreement index. Thus, for the JPI syncritic scores merely consist of proportion of agreement no weights are used.

Subjects and criterion people

All people who participated in this study served both as "subject" and "criterion" group member when syncrisis was performed. Syncrisis involves comparing the responses of two people to the same inventory and assigning a number based on degree of agreement. A large number of syncritic analyses were needed in order to accurately establish correlations. Thus, each person's answers were compared with all remaining peoples' answers. For example, with the KOIS for each subject in the study, all remaining 175 people were rank ordered for degree of similarity to the subject. Referring to Figure 3,
an analogy can be made. If Jones, Wolpe, Wundt, and Freud were all subject and criterion people, as the groups are used in this study, then for each one a ranking of the other three is possible. It might be: Jones-Wolpe, Freud, Wundt; Wolpe-Freud, Jones, Wundt; etc.

Analyses

Ordinarily, test-retest reliability is calculated by administering a test twice, converting subjects' responses to scale scores, and correlating across subjects the scores earned on the first and second test administrations. Syncritic scoring necessitates a somewhat different procedure. It requires that within person correlations be computed for the syncritic score orderings of all "criterion" people compared to each subject at the beginning and end of the three week period, separately for each inventory. The average of these correlations serves as an index of reliability for syncritic analysis. An unanticipated consequence of using the test-retest reliability procedure with syncritic scoring was that two kinds of reliability can be assessed. In each case the term "reliability" refers to stability of the syncritic score orderings.

Open-pool analysis Open-pool analysis is schematically diagrammed in Figure 5. The reasons for naming this procedure open-pool analysis are best understood after the method has
STEP 1

Each subject takes KOIS and JPI at Time One.

STEP 2

Each subject's responses are syncritically compared to all remaining subjects' responses.

Example:

<table>
<thead>
<tr>
<th>KOIS</th>
<th>JPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>650</td>
<td>Subject 121, .95 Subject 122</td>
</tr>
<tr>
<td>633</td>
<td>Subject 15, .91 Subject 35</td>
</tr>
<tr>
<td>600</td>
<td>Subject 68, .87 Subject 23</td>
</tr>
<tr>
<td></td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>.</td>
</tr>
<tr>
<td>350</td>
<td>Subject 200, .80 Subject 13</td>
</tr>
</tbody>
</table>

Note: There is no reason to expect subjects to be rank ordered identically by the two instruments.

STEP 3

Using Time Two Data, repeat Step 2.

STEP 5

Correlate the results of Step 2 with those of Step 4.

Figure 5. Schematic of open-pool reliability calculation procedure
been described. Briefly, the steps involved in calculating open-pool syncritic reliability are as follows: Subjects take the inventory for the first time (time A), and after the three weeks' delay, retake the same inventory (time B). Syncritic scoring is then performed on the time A data. This results in an interpersonal similarity matrix of size \( N \times N \), where \( N \) equals the number of people who took the test twice. Each row of the matrix contains syncritic scores representing the degree of similarity between the subject identified with the row and all remaining subjects in the pool who are each identified with a single column.

The next step is to form an interpersonal similarity matrix on the time B data. After this has been done the two matrices each contain one row (or, column—the matrix is symmetrical) of syncritic scores representing the amount of similarity to the subject identified with the row for all remaining subjects. Each row of the time A matrix is then rank order correlated (with Spearman's rho statistic (Siegel, 1956, pp. 202-213)) with its corresponding row in the time B matrix. This involves the intermediate step of ranking the criterion people for amount of similarity to the subject at both time A and time B.

This procedure is termed open-pool analysis because two sources of unreliability figure into the final, averaged rho. The subjects will, of course, be the first source, but secondly,
because the subjects are also the criterion people, the criterion peoples' changes in their responses will act as another source of variation. The result is a low estimate of syncritic reliability because computation of open-pool reliability takes into account two sources of variance. The term "open-pool" was coined to identify the fact that the criterion peoples' answers are "open" to change.

Open-pool analysis estimates the amount of reliable variance for two interpersonal similarity matrices formed on the same people at two different times. At the present, it seems that open-pool analysis might have potential research applications for measuring stability (or, change) of interpersonal similarity of intact groups over any given period of time. Reliability values at, or near the level obtained with "fixed-pool" analysis (see below) would indicate that the group studied was relatively stable (i.e., that interpersonal similarity had not changed between these administrations). When open-pool reliability is substantially lower than fixed-pool reliability, changes among group members' interpersonal similarity might be the cause of the difference. Just what constitutes "substantially lower" has yet to be worked out. In any case, substantial differences between these reliability indexes might be most expected after some sort of treatment has been applied between the two test administrations.
Fixed-pool analysis. The fixed-pool procedure involves these steps: First, an interpersonal similarity matrix is formed on the time A data using the test takers as both subjects and criterion people. This step is identical to the first step of open-pool analysis. The next step is to form an interpersonal similarity matrix using the subjects' time A responses as the criterion pool and the subjects' time B responses as the subject pool. In this way, the criterion peoples' answers have been "fixed" at the pattern given at time A. Rank order correlations are computed between the two matrices. As with open-pool analysis, the average of these within person correlations then serves as the index of reliability. Fixed-pool analysis thus accounts for only one source of variance between test administrations, namely, changes in the subjects' answers. The criterion peoples' answers do not change.

Fixed-pool analysis for syncrisis is patterned after the reliability studies used for standardized psychological tests. In such studies, one doesn't change item keying between test periods. The items are scored the same way every time the instrument is used. The same will be true for syncritically scored inventories if the procedure is adopted for the uses described in previous chapters: A pool of criterion people will be identified, they will give their answers to the questions on the inventory to be used, and their answers once
given will serve as a more or less permanent set of keys for syncritic scoring. Estimated reliability by the fixed-pool procedure is thus the primary type of reliability of import to the experiment reported in this dissertation.

Results

All correlations reported in this section are significantly greater than zero at far beyond the commonly adopted alpha = .01 level as tested by the formula provided by Siegel (1956, p. 212). Unfortunately, as happens with many non-parametric statistics, there is no widely accepted method for testing the significance of the difference between two rho correlations. Some solution to this problem will need to be arrived at before much use can be made of the fixed-pool, open-pool reliability distinction as outlined above (see Discussion).

The reader should be advised that corrections were not made for tied ranks in this study. The number of calculations required for computing the Spearman rho correlations reported on was quite enormous. Because of the magnitude of the correlation calculation problem, which would have been multiplied if ties had been accounted for, corrections were not made for tied ranks. This procedure was justified by the fact that the effect of not correcting for ties is known to be a quite small inflation of rho (Siegel, 1956, pp. 207-210, where the correction shaves only 0.001 off the uncorrected figure). Other
factors (e.g., change in sample size, use of a different delay period between test administrations, or use of samples with different characteristics than those reported here) could easily have influenced the values of rho obtained for this study to a greater degree than the lack of tie correction. The focus of this study is on whether syncriisis can be used reliably. A modest amount of accuracy was sacrificed, but the general questions of interest were still answerable. For these reasons, it seemed justifiable to omit tie correction from the computations.

Table 1 shows the distributions of individual subjects' rhos for the fixed-pool analyses for both the JPI and the KOIS. Separate tabulations for males and females are also shown in Table 1. All distributions are negatively skewed. Therefore, the median, as opposed to the mean, is perhaps the better figure to use as the index of average reliability. Median reliabilities are often used by test publishers (e.g., Campbell, 1977; Kuder, 1974) when distributions are skewed. In this section, median correlations are reported followed in parentheses by the corresponding mean figure.

The median fixed-pool rho for the KOIS was 0.881 (M = 0.866). Slightly lower were the averages for the JPI, Mdn = 0.797 (M = 0.782). As expected, the open-pool analysis yielded lower estimates of syncriisis reliability. The averages for open-pool reliabilities were Mdn = 0.788 (M = 0.756) for
Table 1. Distributions of individual subjects' rho correlations for fixed-pool reliability analyses

<table>
<thead>
<tr>
<th>Interval of rho</th>
<th>KOIS</th>
<th>JPI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males (N) Females (N) Total</td>
<td>Males (N) Females (N) Total</td>
</tr>
<tr>
<td>1.0-.95</td>
<td>1 0 1</td>
<td>0 0 0</td>
</tr>
<tr>
<td>.90</td>
<td>26 37 63</td>
<td>0 1 1</td>
</tr>
<tr>
<td>.85</td>
<td>18 34 52</td>
<td>7 16 23</td>
</tr>
<tr>
<td>.80</td>
<td>18 14 37</td>
<td>21 36 57</td>
</tr>
<tr>
<td>.75</td>
<td>9 5 14</td>
<td>19 19 38</td>
</tr>
<tr>
<td>.70</td>
<td>4 2 6</td>
<td>14 13 27</td>
</tr>
<tr>
<td>.65</td>
<td>0 0 0</td>
<td>8 4 12</td>
</tr>
<tr>
<td>.60</td>
<td>1 0 1</td>
<td>4 2 6</td>
</tr>
<tr>
<td>.55</td>
<td>0 0 0</td>
<td>0 1 1</td>
</tr>
<tr>
<td>.50</td>
<td>1 0 1</td>
<td>0 1 1</td>
</tr>
<tr>
<td>.45</td>
<td>1 0 1</td>
<td>1 0 1</td>
</tr>
<tr>
<td>&lt;.45</td>
<td>0 0 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Totals</td>
<td>79 97 176</td>
<td>74 93 167</td>
</tr>
</tbody>
</table>
the KOIS, and \( \text{Md}_n = 0.686 \) (\( M = 0.678 \)) for the JPI. Again, the JPI was apparently a less reliable basis for syncritic scoring than was the KOIS. Table 2 shows the distributions of rhos for the open-pool analyses for the JPI and the KOIS.

Differences in reliability between the sexes appear to be quite small. These differences range from 0.07 (JPI open-pool) down to approximately 0.03 (KOIS fixed-pool). Table 3 shows the median test-retest correlations for males and females calculated by fixed-pool and open-pool procedures for both inventories. Even without a method for testing the statistical significance between rho correlations, it appears permissible to conclude that there are no substantial sex differences in this study in the reliability of syncrisis.

Within person Pearson product-moment correlations were computed across scales as estimates of the reliability of JPI profiles. (As noted previously, KOIS profiles were not available to the author.) The average of these correlations served as the estimate of reliability for this sample of subjects when standard scale scoring was used. The median Pearson \( r \) was 0.814 (\( M = 0.807 \)). The distribution of these correlations is shown in Table 4.

The data in Table 4 are of interest because they serve as a standard for judging the performance of syncritic scoring technology. The profile-level reliability data in Table 4 indicate a kind of maximal reliability to be expected when a
Table 2. Distributions of individual subjects' rho correlations for open-pool reliability analyses

<table>
<thead>
<tr>
<th>Interval of rho</th>
<th>KOIS</th>
<th></th>
<th></th>
<th>JPI</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males (N)</td>
<td>Females (N)</td>
<td>Total</td>
<td>Males (N)</td>
<td>Females (N)</td>
<td>Total</td>
</tr>
<tr>
<td>1.0-.95</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>.90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>.85</td>
<td>10</td>
<td>4</td>
<td>14</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>.80</td>
<td>12</td>
<td>30</td>
<td>42</td>
<td>7</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>.75</td>
<td>21</td>
<td>29</td>
<td>50</td>
<td>9</td>
<td>19</td>
<td>28</td>
</tr>
<tr>
<td>.70</td>
<td>13</td>
<td>20</td>
<td>33</td>
<td>11</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>.65</td>
<td>13</td>
<td>9</td>
<td>22</td>
<td>13</td>
<td>12</td>
<td>25</td>
</tr>
<tr>
<td>.60</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>11</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>.55</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>.50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>.45</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>.40</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>&lt;.40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Totals</td>
<td>79</td>
<td>97</td>
<td>176</td>
<td>74</td>
<td>93</td>
<td>167</td>
</tr>
</tbody>
</table>
Table 3. Median reliabilities for males and females

<table>
<thead>
<tr>
<th>Sex</th>
<th>Type of Analysis</th>
<th>Fixed Inventory</th>
<th>Open Inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KOIS</td>
<td>JPI</td>
<td>KOIS</td>
</tr>
<tr>
<td>Male</td>
<td>0.868</td>
<td>0.773</td>
<td>0.777</td>
</tr>
<tr>
<td>Female</td>
<td>0.893</td>
<td>0.841</td>
<td>0.773</td>
</tr>
<tr>
<td>Difference</td>
<td>0.025</td>
<td>0.068</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Table 4. Distribution of JPI profile level reliabilities

<table>
<thead>
<tr>
<th>Interval of Pearson r</th>
<th>Males(N)</th>
<th>Females(N)</th>
<th>Total(N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0-.95</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>.90</td>
<td>15</td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>.85</td>
<td>16</td>
<td>10</td>
<td>26</td>
</tr>
<tr>
<td>.80</td>
<td>17</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>.75</td>
<td>8</td>
<td>20</td>
<td>28</td>
</tr>
<tr>
<td>.70</td>
<td>6</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>.65</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>.60</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>.55</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>.50</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>.45</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&lt;.45</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Totals</td>
<td>74</td>
<td>93</td>
<td>167</td>
</tr>
</tbody>
</table>
purportedly highly reliable (Jackson, 1976) inventory is scored in standard fashion for a test-retest sample. The median correlation for these within person reliabilities was found to be 0.814. Comparatively, the median correlation for JPI fixed-pool analysis was 0.797. Although the former correlation is a Pearson $r$, while the latter is a Spearman rho, it is possible to make comparisons between the two (each was calculated on the basis of its appropriateness to the data which were correlated). The reason comparisons of a judgmental sort may be made between rho and $r$ is that rho serves to estimate what the Pearson $r$ would have been had the data been of the interval instead of the ordinal variety (cf., Nunnally, 1967, p. 121). In fact, had Pearson $r$'s been calculated on the rank order data described in this section exactly the same coefficients would have resulted as found by the Spearman method! Considering the small size of the difference between these two reliability indexes (about 0.02) there seems to be nothing substantial to gain by testing for statistical significance of the difference.

Discussion

Quite obviously, one consequence of proposing to study a new method for psychological assessment was that numerous problems were encountered during the data analysis phase of
the study. Much of the Results section might appear to be statistical nitpicking, but the type of data yielded by syncrisis were difficult to examine with standard statistical techniques. Nonetheless, the major aim of the study, estimation of syncritic scoring reliabilities for two inventories, was met. Furthermore, the obtained reliability estimates were large enough to warrant a fairly optimistic discussion of the implications of the experimental results for future research and practical applications.

Consider first the median fixed-pool reliabilities for the KOIS (rho = 0.866) and the JPI (rho = 0.797). The median rho for the KOIS is only slightly lower than within person reliability coefficients reported in Kuder's (1974, p. 37) KOIS manual. The manual states that a sample of high school and college subjects tested, then retested two weeks later showed a median reliability for rank ordered scale elevation of 0.90. Test-retest data have not been made available for the JPI (Jackson, 1976); reported reliability estimates were arrived at with Cronbach's alpha. However, the experiment designed for this dissertation produced a median within person Pearson r of 0.814 for profile reliability in a test-retest situation. The median profile reliability is rather close in magnitude to the median fixed-pool reliability for the JPI (see above). In view of the relationship between the Pearson r statistic (used to calculate the profile reliability)
and the Spearman rho statistic (used to calculate syncritic fixed-pool reliability), it seems acceptable to conclude that comparatively equal reliabilities were shown for the JPI for both the standard and the syncritic scoring methods.

The finding that syncritic scoring was nearly as reliable as conventional scoring for the two inventories studied gives credence to the claims made in earlier chapters about the potential utility of the new method for psychological assessment. Recall that reliability was one of two criteria a test or scoring method must meet in order to be acceptable for use as an assessment method. Since syncritic reliability was comparable to standard scoring reliability, it follows that the syncritic method meets the criterion of acceptable reliability, at least for a three week test-retest situation.

The brief interval between test administrations used in this study could be lengthened substantially (say, to a year or more) in a future research project. Long term test-retest reliabilities are routinely provided by the publishers of widely used psychological inventories (e.g., Campbell, 1977; Kuder, 1974). If item responses from a long term test-retest study were available, one could proceed as directed in the Methods section to estimate the fixed-pool reliability of syncrisis for these samples. In this way, long term test-retest reliability for syncrisis could be estimated without much delay.
In addition to yielding overall estimates of syncritic scoring reliability, Experiment 1 provided other important data. First, no substantial differences were found in reliability levels attained for males and females. Interpretation of the experimental results, and the possibilities for applying syncrisis, in other situations, would have been markedly complicated if sex differences had emerged. There really was not any reason to expect sex differences to occur, however, quite often in psychological research, differences in performance (or, attitudes, personality, reliability, etc.) for the sexes are observed. The present data permit a tentative conclusion that syncritic analyses may be performed with equal reliability for both males and females.

An intriguing, if not important, "result" was the recognition of the open-pool versus fixed-pool distinction in reliability estimation procedures. Although the validity of applying this distinction as a research methodology has not been empirically tested, there are circumstances for which the method may prove useful (see Analyses subsection). Prior to applying the method as a research tool, however, some problems must be ironed out by further work. One problem, of course, is the matter of determining the statistical significance of a difference between two rho correlations inasmuch as the key to the method lies with finding fixed-pool minus open-pool differences. A step toward solving the problem has
come from consultants at the Iowa State University Statistical Laboratories. According to them, if a rho is based on a large number of observations (40 or more) it may be subjected to the same tests as a Pearson r correlation. This is only a partial solution because the data analyzed for the present study show that there is a difference between fixed-pool and open-pool reliabilities even when no "treatment" intervened between the test-retest inventory administrations (Table 3). This problem may turn out impractical to solve, hence the fixed-pool open-pool distinction may remain intriguing and not important.

To summarize the results of Experiment 1, it appears that syncrisis has met the criterion of showing adequate reliability. The next two chapters contain studies of the validity of syncrisis for the purpose of client-therapist matching. The introduction to the next chapter explains the rationale for client-therapist matching, reviews previous matching research, and describes why syncrisis might be expected to be useful for making therapeutically helpful matches.
EXPERIMENT 2: CLIENT-THERAPIST MATCHING

It was asserted in an earlier chapter that syncritical analyses may, in the first place, represent improved methods for reaching the stated goals of traditional assessment, and that secondly, the technology of syncrisis makes possible the meeting of other, nontraditional aims. To be sure, standard scoring techniques can be applied to objective inventories to fulfill unique purposes, and satisfactory results can be had. It remains possible, however, that for some instances, if not for the majority, standard scoring is less suitable than syncrisis. An instance of applying objective psychological instruments for nontraditional purposes which seems to present such a possibility is that of "matching" therapists and clients as a precursor to psychotherapeutic counseling. What follows is a discussion of the rationale for client-therapist matching, a brief but critical review of the literature on the subject, and a comment on the technologies which have been used to date illustrating why syncritical methods may be superior.

Recall that one goal of psychological assessment has been to facilitate psychological treatment (e.g., Blashfield & Draguns, 1976; Woodruff, Goodwin, & Guze, 1974). Historically, the assessment effort has been directed at the first half of the client-counselor dyad. The logic of this approach begins with the clearly defensible premise, analogous to the one used in medicine, that different disorders (or, problems, or people
with problems, etc.) require different treatments (cf., Millon, 1969; Phillips & Draguns, 1971). Thus, assessment of the clients' personality or behavioral characteristics (not itself unproblematic see the earlier discussion of diagnosis) should aid treatment since the therapist could make adjustments in the counseling process to maximize the effects in the problem area. But, in the context of influential, eponymic figures' (e.g., Rogers, 1957; Freud, 1949) assertions that single, standardized counseling approaches were best, early research cast the outcome effectiveness questions as, "does psychotherapy work?" (Bergin, 1971; Bergin & Lambert, 1971). Subsequent criticism (Kiesler, 1966) that such research prompted a "uniformity myth" viz, that clients and therapists are each homogeneous populations, prompted researchers to investigate the relative effectiveness of various kinds of therapy on people of specific diagnostic types, or with specific anomalies (Paul, 1967; Sloanne, Staples, Cristel, Yorkston, & Whipple, 1975; DiLorete, 1971; Lazarus, 1976). Also following Kiesler's criticisms, has been increased emphasis on investigations of the second half of the client counselor dyad, that is, study of the therapist variables, such as personality characteristics, experience level, sex, etc. on outcome (cf., Parloff, Vaskow & Wolfe, 1978).

Even before the recognition of the importance of avoiding uniformity myths, there were early studies of successful
versus unsuccessful therapists (Whitehorn & Betz, 1957, 1960). Such research constitutes the acknowledged first instance of client-therapist matching. The tacit assumption of all matching research is that important features of counselors' and clients' personalities will influence the process and outcome of therapy. It is further assumed that, at least on some dimensions, similarity on these dimensions will facilitate the client's progress in therapy. Two studies designed for this dissertation, for example, assume that congruence of interest patterns might be helpful for clients (Experiments 2 and 3).

Admittedly, at present there is little in the various theories of psychotherapy which suggests such matching will offer any benefit to the client. Rather, this sort of research has arisen out of empirical and pragmatic considerations.

People have long recognized individual differences among themselves, it is known that some counselors are more effective than others, and some types of clients seem to benefit more from therapy. From awareness of such facts and a concern for the need to be specific about what the client and therapist were like in research studies, some psychologists have been led to the notion that perhaps similarity between the dyad members is a key ingredient in a beneficial counseling relationship. Furthermore, interest in matching clients with therapists has been shown by mental health agency administrators (Berzin, 1977). Such persons direct a large pool of
therapists who in turn conduct therapy with an even larger number of clients. From this perspective, pragmatic issues such as increasing benefits to clients, as well as efficient use of agency resources become important. In fact, Berzin has conducted a series of studies on matching in a mental health clinic considered to be the most extensive yet performed.

Matching Research

Clients and therapists may be matched on any of a variety of bases. The simplest method might use demographic variables such as sex, race, or age. Parloff et al. (1978) in a recent review of studies employing demographic matching variables have concluded that there is little use in such research. In addition to methodological problems, there is seemingly little relationship between biological factors (sex, race) on the one hand and a psychological process (therapy) on the other. In the Parloff et al. view, it would be more useful to study the psychological correlates of demographic variables (e.g., therapists' attitudes toward other and own sex) as mediators of effects attributed to demographic variables.

Two other potential sets of dimensions for making matches are activity interests and personality traits. These are the dimensions for matching employed in the present research, and accordingly, below are reviewed only studies which focus upon
one or another of these avenues for determining client-therapist matching (CTM). However, since the major aim of this thesis is to study syncritic analysis (its use in CTM is considered a test of validity), I have not attempted to be exhaustive in the discussion below.

Although Whitehorn and Betz (1954, 1957, 1960) are credited with initiating the client-therapist matching research, their reports actually consist of analyses of differences between A-type (successful with schizophrenics) and B-type (not successful with schizophrenics) therapists in a psychiatric hospital setting. Thus, in the sense of matching (viz, similarity of client and counselor) as I have used it, they did not investigate matching. Nevertheless, their 1954 publication begins with the assertion that, "It is difficult for one physician to plan and maintain crucially different approaches and attitudes and patterns of interaction . . ." (p. 321). This comment that any one man or woman as therapist may not be capable of making the kinds of personal, stylistic adjustments to suit every kind of client suggest that Whitehorn and Betz were sensitive to the issues which later, through the efforts of other researchers, came into being as client-therapist matching. For this reason, below are reviewed the highlights of the three studies they produced.

The 1954 study was based on data provided by 14 psychiatric residents at Johns Hopkins Hospital who treated a total
of 100 schizophrenic people. The 14 physicians were selected from an initial group of 35, all of whom had treated at least 4 each of schizophrenic, depressed, and neurotic patients. Seven of the 14 were the most successful of the 35 schizophrenics (75% average of their patients were discharged as "improved"), while the other seven were the least successful of the 35 physicians (27% average discharge rate). The authors then rated the case records of the 52 patients who had been treated by the A-type therapists and the records of the 48 B-type therapists' patients for five "characteristics", which were types of outcome, relationship quality, diagnostic formulation, goals, and "tactical patterns". Based on unnamed statistical analyses, Whithorn and Betz were able to conclude that A-type therapists were more active, aimed at "assisting the patient in definite modification of personal adjustment patterns and toward more constructive use of assets rather than mere decrease of symptoms . . ." (p. 331), and showed "some grasp of the personal meaning and motivation of the patient's behavior" (p. 331). There were no differences in effectiveness with depressives and neurotics. Their 1956 and 1960 publications reported the results of their use of the Strong Vocational Interest Inventory, Form M (Strong, 1943) to differentiate between the A and B therapists. They chose to use the SVIB-M because it measured interests instead of psychopathology. Based on the SVIB-M profiles of the 14, 1954
therapists, and 12, 1956 subjects they found that a combination of four scales differentiated between A and B types: A's high on CPA, and lawyer, low on Printer and Mathematics. They cross validated their "rationally"-developed group inclusion rule on a further sample of 24 residents. (The inclusion rule assigned 0 - 4 points for agreement with the A-type pattern, i.e., 4 points for high CPA and Lawyer, low Printer and Teacher). Predicting A-types for scores of 3 or 4 resulted in "correct" classification of 8/10 A-therapists. The authors permitted themselves to conclude that "A's, with interests resembling lawyers have a problem-solving, not purely regulative of coercive approach. This is acceptable to the resentful, boxed-in patient . . . " (p. 222).

These and other A-B therapist studies have been subjected to heavy criticism, especially by May (1974) who reanalyzed the Whitehorn-Betz data and came to skeptical conclusions. By contrast, Razin (1971, 1977) began as an optimist. But in 1977 he wrote, "My involvement with the A-B variable has been marked by initial optimism, as reflected in the 1971 review; this optimism has progressively diminished and turned to skepticism" (p. 319). Over 100 published articles, 18 different A-B "scales" (Stephens, Shaffer, Zlowtowitz, 1975) and numerous reviews have been produced on the subject. The A-B dimension research seems to have been a fad (cf., Meehl, 1978),
one that appears to be headed for the fate of the California-F scale. It was, however, the stimulus for more direct study of therapist and client variables in combination.

Matching research is presently proceeding along several avenues. There are two recently published reviews of the subject, the Parloff et al. chapter in Garfield and Bergin (1978) and three chapters covering 100 pages in Gurman and Razin (1977). The rest of this section on CTM consists of a summary of the findings presented in these two reviews, a discussion of one research program in detail, followed by the rationale for using syncrisis, especially with activity interests.

The Parloff et al. (1978) review covers only studies reporting results from clinical investigations; there is no mention of analogue study results. The dependent variable typically was outcome or success. They noted (p. 265) that the studies they reviewed focused on three factors in CTM research: personality, cognitive, and values. Evidently, then, there has been little emphasis (I found none) on interest pattern compatibility, a point to which the discussion returns below.

Personality similarity can produce some unusual results. The Parloff group cites Wogan's (1970) study as an example. He had used the MMPI as a means for CTM, and Parloff et al. misleadingly conclude that, "he found that therapist patient similarity was a detriment to therapy" (p. 266). What is misleading is that the MMPI measures pathology, thus it makes a
good deal of sense that similarity in this regard was detri-
mental. Wogan had canonically correlated MMPI scale scores
(without correcting for overlap, it should be added) and
identified five "factor scales". Two of these were named
Repression and Subtlety and these two were the ones for which
CTM contributed negatively to the dependent variable pre-
dictibility, as evidenced by a significant contribution to a
stepwise multiple-R by the therapist and patient cross-products
on these "factors". Interestingly, dissimilarity on the fac-
tors contributed nothing to patient and therapist ratings,
the dependent variables.

Doughtery (1976) also obtained interesting results when
he matched clients to therapists on the basis of personality
variables. He used multiple regression to form predictive
equations on a derivation sample of therapists and clients
who had all responded to 11 scales from a variety of psy-
chological inventories. The criterion predicted was therap-
ist rating of outcome. The derived equations were then used
to assign patients to "optimal" or "deterioration" match
therapists. The main finding was that optimal matching did
not result in greater therapist-rated improvement than random
assignment, but deterioration matched clients were given sig-
nificantly lower ratings compared to controls. Apparently,
then, similarity between client and therapist had no effect
while dissimilarity resulted in significantly worse treatment
"benefits" for clients.

One way to overcome the problems of matching on personality variables is to use inventories that are not pathologically oriented. This, however, is not a total solution since it is probable that for some personality dimensions CTM might be optimized if the client and therapist were different. Gassner's (1970) dissertation study recognized these possibilities. She assessed therapists (12 theology students) and clients (hospital inpatients) with the Fundamental Interpersonal Relations Orientation (FIRO-B) scale, which purportedly measures compatibility as a function of similarity and dissimilarity. Each of 12 therapists treated two compatible and two incompatible patients. A t-test revealed that patients in compatible dyads rated their therapist much more favorably than those in incompatible dyads. However, neither the high nor low compatibility dyads improved, as evidenced by pre- and posttreatment ratings by ward nurses (a somewhat problematic criterion, see Garfield and Bergin, 1978, p. 250).

Gassner developed her study out of the social-influence literature (see Strong, 1978) which postulates that therapy is a form of interpersonal influence, and relies heavily on findings that interpersonal liking increases susceptibility to interpersonal influence. These notions provide a backdrop for understanding part of the rationale for Experiment 2. Specifically, the study is concerned with two questions. One, is
syncritic analysis of clients and therapists inventory responses of potential use in CTM? That is, CTM research itself is study of a sort of syncrisis (in the sense of individual comparison), so it becomes relevant to ask if it can be improved when more thorough individual comparisons were made. Two, are communality of interest patterns of perhaps greater import than global similarity on personality dimensions? That is, since the goal is to enhance interpersonal liking (from the social psychology point of view), then it seems that congruent interests could be significant in increasing liking. Experiment 2, described in the next section, was designed to provide partial answers to these two questions. Before completing the present section, a discussion of Berzins' (1977) Indiana Matching Project (IMP) is in order because it supports the idea that CTM optimizing has some promise in improving clients' benefits from therapy.

IMP was a four-year long research effort aimed at developing a CTM algorithm at a university psychology clinic where the emphasis was on brief therapy (four weeks or less). The project had two phases. During the first phase, the ten project therapists were tested with the Personality Research Form (PRF; Jackson, 1967). The therapists' PRF scores were converted to factor scores from a previously performed factor analysis of 156 PRF profiles. All therapists were then categorized into homogeneous personality groups which were used
as a "type of therapist" factor for an analysis of variance experiment. Clients (n = 751) were categorized into a "type of client" factor on the basis of their answers to a pretest. The dependent variable was client improvement, jointly rated by both members of each therapy dyad. When the ANOVA was computed the first phase of the study had ended. Its purpose was to identify statistically significant client x therapist interaction effects to be used in prediction equations for phase two of the study.

Phase two of Berzin's IMP proceeded along these lines: Incoming clients filled out the pretest, and their scores were used in the prediction equations resulting in one of three outcome predictions for each of the ten therapists. These were, "assign"—an optimal match, "assign if necessary"—a moderately optimal match, and "don't assign"—a deleterious outcome was predicted for this group. Control groups were formed by random matching, members of one control group were given "placebo compatibility instructions", that is, they were told they had been matched to a therapist when, in fact, the assignment was random. The results of this validity study were as follows: 1) Bogus matches between clients and counselors resulted in no better improvement than random matching. Thus, placebo effects were ruled out. 2) Optimal compatibility led to significantly greater improvement in clients compared to improvement in the other two matching groups combined.
CTM seems to have potential as a means for increasing the effectiveness of mental health service delivery. The IMP study demonstrates this. The problem with Berlin's IMP study, and Doughtery's (1976) experiment as well, is that it is quite unlikely that their predictive equations would cross-validate. In both studies, therapists in the predictive equation derivation samples were used in the validation experiments. What is needed is a methodology more general in scope than these ad hoc procedures. CTM is a form of assessment in that results of pretreatment inventories given to clients are used to make decisions about therapy for these people. As a form of assessment, CTM might be improved through the use of syncritic psychometrics. Experiment 2 explores some possibilities for improving CTM through providing predictions about clients' liking ratings of potential therapists.

Method

Sample

Six graduate students matriculated in the Counseling Psychology program at Iowa State University served as therapists for this study. There were three male, and three female therapists. They agreed to participate after being asked to do so.

The term "client" in the context of Experiment 2 refers to undergraduate subjects who volunteered to participate in
exchange for course credit. There were 52 clients (25 male, 27 female) who provided data for this study.

Procedures

Therapists were given a set of questions about their personalities, their interests, and their counseling styles (Appendix A). The questions were chosen to cover in a balanced fashion both interests and personality factors. This balance was attempted because syncritic predictions were made on the basis of clients' and therapists' responses to the JPI and the KOIS (described in Experiment 1).

Each therapist was given one week to review the questions. Therapists reported individually at appointed times to a laboratory room where they were asked to provide frank, confidential answers to the personal questions. These answers were tape recorded while each therapist was alone. They were told that their answers would serve as stimulus materials in another experiment (the nature of which was not described). They were further instructed that there would be no identification attached to their answers and that no one except the present author and the experimental subjects would have access to the stimulus materials they provided. Therapists responded to the JPI and the KOIS after audio recording their answers to the projective questions shown in Appendix A. Following this, they were told that the purpose of the study was to investigate the relationship between similarity of therapists' and clients'
inventory responses on the one hand, and clients' preferences for seeing specific therapists, on the other. Questions about the study were answered after the study was described.

Therapists' answers to the three open-ended questions were transcribed and typed up in a standard format. This resulted in six, three page, double-spaced transcripts which had three sectional headings. The headings were added as a convenience to the clients, each heading was written as a question and referred to personality, interest, or counseling style content, in that order. A minimum of editing of the tapes was required because instructions to the therapists clearly specified the format of the transcripts the author intended to show to the clients. Each transcript began with the heading "Counselor Number __", the blank contained a number from 1-6.

Clients were seen in groups of two to six people. Instructions were given in two parts, corresponding to the two-part nature of the client's task. For the first task the clients were simply told that the study was concerned with the relationship between peoples' answers to "psychological questionnaires" and their choices for working with preferred counselors, if they were a person with a problem. (Examples of problems were suggested as roommate conflict, academic problem, career choice indecisiveness.) They were further told that there were two parts to the study. Part one began
right after the instructions and consisted of taking the JPI and KOIS, each of which was briefly explained. Part two was the counselor rating task, not further described until the inventories had been completed.

The instructions for the rating task were written on the response sheet. A facsimile is shown in Appendix B. The instructions were, in brief, to read through the six therapists' transcripts and then rank order the counselors from most to least preferred for seeing each of them if the client were a person seeking counseling. A seven-point Likert item for amount of preference for seeing each counselor was also given to the clients. Client-subjects required a little over two hours to complete the experimental tasks.

**Analyses**

All six of the therapists were syncritically compared to each client resulting in two sets of syncritic score ranks of all six therapists for each of the 52 subjects. One set of ranks was based on the KOIS, the second set was based on the JPI. A third set of rankings of each therapist relative to each client was obtained by scoring the JPI in standard fashion and correlating therapists' and clients' profiles with a Pearson-\(r\). The resultant coefficients, six per client, were then ranked from the largest to smallest in value. This third set of rank orderings is referred to as "profile-level ranks" in the Results section below.
RESULTS

The primary finding of Experiment 2 was that clients' preferences for counselors were not predictable under the circumstances studied. Mean Spearman rho correlations between the syncritic ranks of therapists compared to clients and the dependent variable of clients' self-reported ranked preferences for clients were not significant. For the KOIS the mean rho was 0.151, NS, and for the JPI the mean rho was 0.029, NS. The mean rho between JPI profile-level ranks and the clients' self-reported ranked preferences for counselors was also not significantly different from zero (M = -0.015). Table 5 shows the distributions of the within person correlations between these three predictor variables and the criterion variable of clients' self-reported ranking of therapists. Each distribution shows a central tendency of approximately zero. There is little skew shown by these distributions, thus the mean is the appropriate index of the average performance for these predictors.

A potential explanation for the failure of this experiment was deduced from analyses of the data in Table 6 and the Likert-item ratings of amount of preference for each therapist. Table 6 shows the frequency of clients' self-reported preferences for each of the six therapists. The preference ranking data were dichotomized into the number of top-half (rank of one, two, or three) and bottom-half (rank of four,
Table 5. Distributions of within person correlations between predictors and the criterion of client self-reported ranking of therapists

<table>
<thead>
<tr>
<th>Magnitude of Rho</th>
<th>KOIS-Syncritic</th>
<th>JPI-Syncritic</th>
<th>JPI-Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.67 to 1.0</td>
<td>7</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>0.33 to 0.67</td>
<td>8</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>0.0 to 0.33</td>
<td>16</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>-0.33 to 0.0</td>
<td>15</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>-0.67 to -0.33</td>
<td>5</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>-1.0 to -0.67</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Totals</td>
<td>52</td>
<td>52</td>
<td>52</td>
</tr>
</tbody>
</table>

Table 6. Frequency of top and bottom ranks for each therapist

<table>
<thead>
<tr>
<th>Rank</th>
<th>Therapist Number</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Top Half (1,2, or 3)</td>
<td>24</td>
<td>19</td>
</tr>
<tr>
<td>Bottom Half (4,5, or 6)</td>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>Totals</td>
<td>52</td>
<td>52</td>
</tr>
</tbody>
</table>
five, or six) preferences expressed by the total sample of 52 subjects for each of the six therapists. The chi-square statistic (df = 5) was 15.21, p < .01. Apparently, the subjects in this experiment showed strong preferences for some therapists (e.g., number four) while "rejecting" other therapists (e.g., number two). This difference in preferability of the therapists may have been sufficient to override any preference for counselors based on similarity between the clients and therapists.

A one-way F-test on the Likert data supports the hypothesis that this sample of clients showed a consensus of strong preference for some therapists and a corresponding aversion to other therapists. With therapists as the independent variable, and amount of preference as the dependent variable, the analysis showed F = 4.26 (df = 5, p < .001). Evidently, something about some therapists' transcripts appears to have resulted in sweeping favor, or disfavor, with this group of clients.

Discussion

Experiment 2 was designed to determine whether clients' preferences for counselors could be predicted by syncritic technology. The experiment failed to show that this end could be achieved. This Discussion of Experiment 2 centers on two substantive issues, (a) the adequacy of the experimental design for detecting predictable preferences, and (b) the question
of whether the failure of Experiment 2 implies that syncritic technology cannot be usefully employed for the more general purpose of client-therapist matching. Since there are sound reasons for a negative answer to the question posed in (b), it is my sad task to find fault with Experiment 2.

One problem with the experimental design was its lack of statistical power to detect relationships among the variables studied. With only six therapists' transcripts available for clients to rate, a rho of over 0.82 (Siegel, 1956, Appendix Table P) was required for significance at the .05 level. This number of therapists was decided upon in order to make the task easily manageable for the clients. The fact that it is difficult to obtain a significant correlation when small numbers of subjects (here, therapists) are employed may mean that the rating task in Experiment 2 cannot be both manageable for the clients and statistically sensitive.

The other major problem with this experiment was pointed out in the Results section. There was a strong, between subject consensus that certain therapists were preferable. In fairness to the therapists who supplied the transcripts, it must be stated that the consensus may indicate that the client sample was homogeneous with respect to attitudes about therapist valuability, and hence, preferability. Thus, if this conjecture is correct, a different homogeneous sample would have preferred a different subset of therapists.
Furthermore, if a highly heterogeneous sample had been studied, the statistical consequence may have been that a systematic relationship between the predictors and the preference ranks would have emerged. This could be tested empirically with the use of a widely diverse sample, but the problem of the relationship between statistical power and manageability of the task would have to be circumvented.

In sum, this study had severe shortcomings which restrict generalizing the negative findings. In view of this fact, the question of whether clients' preferences for counselors are predictable, by syncritic or other means, remains unanswered.

This study was proposed as an analogue of CTM as it might occur under real circumstances outside the laboratory. Even without the design difficulties cited above, a failure of Experiment 2 may not have implied that syncritic technology could not be useful for CTM purposes. This conclusion arises from the fact that the study reported here used the dependent measure of clients' preference rankings of therapists as an analogue to the dependent variable of interest when CTM is applied in actual counseling situations, viz, better therapy for clients. Experiment 2 could not directly address the latter issue. Had positive results been found, they would have served as empirical evidence that actual CTM might be furthered through the use of the methods employed in this
experiment. Barring empirically supported motives for pursuing the application of these methods to CTM in applied settings, there remain logical reasons for further study of the possibilities. The logic of applying syncrisis to CTM was outlined in the introductory remarks of this chapter. Briefly in review, CTM involves individual comparisons, as the term "matching" denotes. Syncrisis results in a numerical index of the degree of match between individuals, taken two at a time, who have responded to a set of questions. If the content of the set of questions is relevant to the purpose of CTM, then it follows that syncrisis should be useful for CTM.

An example of item content previously found to be relevant to effective CTM are personality characteristics (see beginning of the chapter). Another potentially relevant content domain may be that of interest patterns. In Experiment 2, neither domain of content was useful for predicting clients' preferences for counselors. Additionally, profile-based predictions were also inaccurate. Experiment 3, reported in the next chapter, was a further attempt to apply methods similar to those used in Experiment 2 for the purpose of testing the applicability of syncrisis to CTM.
EXPERIMENT 3: OUTCOME OF ACTUAL COUNSELING AS A FUNCTION OF CLIENT- THERAPIST SIMILARITY

In its most rudimentary form the CTM hypothesis simply holds that, when the client and therapist are similar, more effective counseling will result than when the dyad members are dissimilar. Experiment 2, described in the previous chapter, was an analogue study and thus involved additional intermediate assumptions about CTM (e.g., that clients' preferences for counselors is related to counseling outcome). That experiment failed, although the reasons for the failure may have been unrelated to the soundness of the CTM hypothesis. The study reported in the present chapter was based on data collected from actual counseling dyads. Experiment 3 was a correlational study of the relationship between client-therapist similarity and two forms of outcome variables. This study was based upon the rudimentary form of the CTM hypothesis stated above. It was expected that syncritic measurements of similarity would significantly correlate with the outcome criteria. Also germane to this study was the degree of relationship between two profile-based indices of client-therapist similarity (Pearson r, and D^2) and the outcome criteria. This permitted examination of the relative effectiveness of syncrisis compared to profile-based similarity measures.
Before describing the methods and results used in this study, a discussion of therapy outcome criteria must be considered. When proponents of CTM assert that similarity between client and therapist will lead to more effective therapy, they have in mind a client-oriented definition of effective. Specifically, since therapy is primarily intended to benefit the client, effective counseling outcome means something like, "the client changed for the better." Following this definition quite strictly, clients' benefits, and hence outcome, were not assessed in this experiment. Instead, the evaluations of outcome used for this study were ratings of the therapists' performances at the task of counseling. The reason for this seemingly anomalous circumstance has to do with the situation from which the data for this study were collected: Counselors and clients were graduate-level trainees.

Under these circumstances, it was, of course, quite reasonable to assess the skills of the therapist instead of the changes shown by the client. Although this represents a departure from the usual procedures for measuring therapeutic outcome, it can be argued that, while some caution should be exercised in drawing conclusions from the study, irreparable harm has not been done. In fact, it would be paradoxical to argue that assessing the therapists' performance was not a valid outcome criterion. If effective
therapy implies that clients benefited, then ipso facto the therapists performed well. Accordingly, two measures (one supplied by the client, the other by the course instructor) of therapist performance were treated as outcome measures for this study.

Method

Sample

The sample of subjects for this study were graduate students enrolled in a counseling prepracticum course at a metropolitan university. There were 24 people enrolled in the course, but only those who voluntarily completed the JPI and the KOIS (see Methods section of Experiment 1 for a description of these instruments) served as subjects for the analyses here reported. Both members of 21 counseling dyads (see below) completed the KOIS; 16 client-counselor pairs took the JPI. These inventories were filled out at the beginning of the semester.

Procedure

Counseling dyads were formed by randomly pairing students in the class with the constraint that no two students serve as client and counselor for each other. Each client received six, one-hour counseling sessions. These sessions were held in addition to the weekly course meetings. Counseling was done in clinical laboratory rooms, and sessions were
tape recorded. Clients were instructed to deal with real, nontrivial personal problems during therapy. Therapists were periodically seen by the course instructor for supervision.

**Predictive methods and outcome variables**

There were four measures of client-therapist similarity which were rank order correlated (Spearman rho) with two basic outcome measures. A third outcome measure was formed as a composite of the two basic measures and it, in turn, was also correlated with the similarity indexes.

The four similarity measures were, (a) JPI-based syncritic scores (syncritic scores were explained in Experiment 1—they indicate amount of agreement between two inventory respondents, here the client and counselor), (b) KOIS-based syncritic scores, (c) Pearson r correlation between each client's JPI profile and that of his or her therapist, (d) D² measure between each client's JPI profile and that of his or her therapist.

The outcome criteria came from two sources. One was the clients' ratings of their therapists on the Client Personal Reaction Questionnaire (CPHQ; Ashby, Ford, Guerney & Guerney, 1957). As used in this study, the CPRQ has 33 questions about the therapist, each of which is scored in interval units on a five point scale. The differences between the CPRQ used here and the original developed by Ashby et al., were that some items were deleted and language has been
updated (e.g., using sex-neutral terminology instead of referring to the therapist as "he"). Total score on the CPRQ scale reflects the degree of client satisfaction with the therapist.

The second outcome measure was the instructor's evaluation of each therapist's performance during counseling. The audio tapes served as the basis for making this evaluation, which was one of several factors used in determining students' total course grades. The counseling performance grade, hereafter called "grade", was scaled from one (C) to seven (A). Pluses and minuses in the grade criterion were thus taken into account.

The third outcome criterion was a composite of the CPRQ and grade measures. Many authorities on outcome measurement for psychotherapy have bemoaned the lack of agreement between outcome measures often found in psychotherapy research (e.g., Garfield & Bergin, 1978, chp. 5; Gurman & Razin, 1977, chps. 19 & 21). On the assumption that, "the truth may lie in the between", the predictability of a composite was included in the analyses. The composite was formed by converting grades and CPRQ scores to z-scores based on their referent distributions. Each therapists z-scores on these criteria were then summed to yield a single index.
Results

Table 7 shows the correlations between the four predictive methods and the three criterion, or outcome, measures. The correlations were computed with Spearman's rho. The computations were done as follows: Members of each dyad were assessed for amount of similarity indicated by each of the four similarity measures. Four rank ordered distributions were thus formed. Dyads were also rank ordered on the basis of therapist's grade, CPRQ score, and combined grade/CPRQ score. This resulted in three rank ordered distributions. Each of the four similarity distributions was then correlated with each of the three outcome distributions.

Table 7. Rank order correlations between similarity measures and outcome measures

<table>
<thead>
<tr>
<th>Similarity Measure</th>
<th>N</th>
<th>Grade</th>
<th>CPRQ</th>
<th>Combined (see text)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOIS-syncritic</td>
<td>21</td>
<td>0.416*</td>
<td>0.238</td>
<td>0.459*</td>
</tr>
<tr>
<td>JPI-syncritic</td>
<td>16</td>
<td>0.052</td>
<td>0.127</td>
<td>0.116</td>
</tr>
<tr>
<td>JPI-profile, Pearson ρ</td>
<td>16</td>
<td>-0.514*</td>
<td>0.315</td>
<td>-0.180</td>
</tr>
<tr>
<td>JPI-profile, D²</td>
<td>16</td>
<td>-0.119</td>
<td>0.091</td>
<td>0.095</td>
</tr>
</tbody>
</table>

*p < .05, 1 tail.
Only KOIS-based syncritic similarity produced a substantial, positive correlation with outcome criteria. Interestingly, the significant correlations were with the grade and combined criteria. No method of similarity was significantly associated with clients' CPRQ assessment of their therapists.

A statistically significant, but negative, correlation was found between the JPI profile-based, Pearson $r$ measure of similarity and the grade outcome variable. The $D^2$ measure of similarity did not show any significant correlations with outcome criteria. Also not significant were the correlations between JPI-based syncritic similarity and the three outcome variables.

Evidently there was a moderately positive relationship between the two basic outcome measures (grade, and CPRQ). The correlation between the two was 0.393 ($p < .05$, 1 tail).

Discussion

The results of this experiment support the claim that syncrisis can be usefully applied in client-therapist matching situations. This study was too small to permit wide ranging generalizations based on these data. But, in spite of its shortcomings, degree of syncritically measured similarity was found to be significantly related to two outcome variables while profile-based similarity was not positively
correlated with any outcome measure. Granted that one ought to be cautious when analyzing the results of an exploratory study such as this one, the data obtained nonetheless provide leads to numerous future research possibilities. For this reason, the discussion below is a speculative interpretation of the results of Experiment 3.

Consider first the positive correlations between syn-critically measured similarity based on the KOIS and two of the outcome criteria. One significant correlation was with the instructor's grading of the therapists' performances. A potential explanation of this is that, when novice therapists are confronted with a client widely different from themselves in activity interests, it is difficult for them to adjust their counseling style in such a manner as to allow them to appropriately counsel the dissimilar client. Conversely, when counseling dyad members are similar, the therapist may have an easier time counseling the client thus creating a situation where the therapist performs at his or her best. This analysis merits researching since, if the conjecture turns out to be correct, it has rather obvious implications for training (and grading) neophyte counselors.

Contrary to the findings with the KOIS, JPI-based syn-critic similarity did not correlate with any of the outcome variables. The discrepancy between the results of using syn-crisis with the KOIS and the JPI can be reconciled in a
manner that suggests testable hypotheses. For example, it may be that interests are more relevant as a content domain than are personality characteristics. This can be tested by simply using multiple measures of personality and interests in a CTM experiment. An additional variable to study in a future research project is the impact of item formats on making syncritic predictions. Recall that the KOIS is in effect a multiple-choice inventory, while the JPI is a true-false test. It is possible that the multiple-choice format, which allows for varying degrees of similarity to be assessed for each item, is more sound as a basis for syncritic similarity measurement.

Another substantively interesting finding was the negative relationship between the grade criterion and JPI profile-based similarity as measured with Pearson's $r$ correlation. This finding was in agreement with Gassner's (1970) earlier cited hypothesis that better outcome may result when the client and the therapist are different. However, when the JPI was used syncritically, no correlation was found with the grade variable. Gassner's matching through dissimilarity idea appears to be method-specific. In any case, the difference in results obtained on the same set of questions (the JPI) when syncrisis as opposed to profiles were the basis for assessing similarity indicates that the two methods are not equivalent. It remains for future, better designed,
research projects to determine which method is best for which applications.

The data on the CPRQ outcome variable present interpretive difficulties. On the one hand, CPRQ was significantly related to the grade measure, indicating some congruence between the clients' and the instructor's evaluations of the therapists' performances. On the other hand, CPRQ was not related to KOIS-based syncritic similarity. Concurrently, KOIS-based syncritic similarity was related to the grade variable. Thus, two outcome variables were significantly related to each other, but only one of these, the grade, was related to a similarity measure. Perhaps the best way to reconcile this arises from recognizing the multifaceted nature of counseling outcome (cf., Garfield & Bergin, 1978, chp. 5). Apparently, syncritically calculated similarity based on the KOIS was related to that aspect of "outcome" which was reflected by an outside observer's evaluation of therapist's performance. The predictability of various facets of outcome by various means provides further research possibilities.

As it turned out, the only significant correlation with the combined CPRQ/grade criterion was from the KOIS-based syncritic similarity measure. In view of the fact that the latter measure was significantly correlated with the grade variable, which was one of two components of the combined criterion, this finding was not surprising. The combined
criterion was formed for the purpose of finding out if such a joint variable might correlate with the similarity measures even if the two component variables were not predictable. This did not happen. Still, it seems possible that under other circumstances, composite outcome criteria may prove useful as dependent variables since "outcome" is a multi-dimensional entity.

It cannot be overstated that the results of this study are tentative. Experiment 3 was included in this dissertation as an illustration of how syncritic technology can be applied to problems presently attacked by other methods. Syncrisis is a new method for measuring interpersonal similarity. Accordingly, the present Discussion has suggested possible applications of syncritic methodology in the context of research designed to test the utility of this method compared to standard methods.
CONCLUSIONS

The objectives of this dissertation were to examine some of the psychometric properties of syncritic scoring technology in studies of reliability and validity. This was done with two inventories, the KOIS and the JPI. It was possible to score the latter in standard fashion, thus permitting a relative benchmark for evaluating the consequences of syncritic scoring.

In review, syncritic scoring was found to be about equally reliable to standard scoring in a test-retest study. Both standard and syncritic scoring reliabilities (approximately 0.80) were high for the JPI. The reliability of syncritic scoring of the KOIS (approximately 0.87) was quite close to figures published in the KOIS manual (Kuder, 1974) which were based on standard scoring of that instrument. The first validity study (Experiment 2) showed that neither syncrisis nor standard scoring was capable of producing a similarity estimate which was, in turn, correlated significantly with clients' self-reported preferences for counselors. The second validity study (Experiment 3) produced data which indicate that syncrisis merits further study as a methodology for accomplishing client-therapist matching.

Taken together these three studies indicate that the reliability of syncrisis is sufficient for applied uses, but the question of the validity of the method remains open.
Experiment 2, for example, was apparently inadequate as a validity test inasmuch as no method of similarity measurement studied was capable of predicting the criterion. Experiment 3 produced positive results for syncrisis, but that study was primarily illustrational, since it lacked adequate controls and was based on a small sample. Obviously, further study of the validity of syncrisis is required.

Throughout this dissertation, I have identified possible research and applied uses of syncrisis. This was appropriate because I wanted to encourage the development of this method. Up to this point, the reader has been exposed to an optimistic, even persuasive, often speculative discussion of syncrisis. My enthusiasm for syncrisis, in light of the results reported in this dissertation, seems warranted. Yet, it must be recognized that little criticism of the new method has been articulated. In what follows below, I have attempted to balance this account of syncrisis. Three major criticisms of syncrisis are described in separate sections. The three were selected for inclusion in this discussion because they represent the major methodological and conceptual problems evident at the present time. If the treatment of each is brief, it is because I don’t have well formulated responses for these objections.
Criticism 1: Syncrisis Produces Impoverished Data

The first criticism arises from the fact that syncrisis results in ordinal data. Furthermore, comparison of one person (e.g., a client) to numerous others (e.g., a criterion pool) involves the undesirable consequence that the resultant syncritic scores are ipsative. The latter problem occurs as follows: If the criterion group is \( N \) in number, then it is clear that an interpersonal similarity matrix of size \( N \times N \) may be formed on the criterion group. This matrix would show that there are varying degrees of similarity among the criterion people. As a consequence, any client's similarity to one criterion person will automatically result in similarity to certain other people as well as dissimilarity to certain others.

My solution to these problems was to analyze syncritic data only with nonparametric statistics, which took into account the ordinal nature of the data. This meant the loss of some statistical power. There is some empirical evidence (e.g., Borgen & Scott, Note 3) that a syncritically formed interpersonal similarity scores can be successfully used in parametric statistics. The Borgen and Scott study showed that valid taxonomic groups could be formed by cluster analysis of a syncritically formed interpersonal similarity matrix (containing syncritic scores). The clustering technique treated absolute values of the syncritic scores, as opposed to first
converting scores to ranks as the data analyses for this dissertation were done. If further studies demonstrate that syncritic scores may be used directly, without ranking them, then possibly this criticism will prove to be pointless.

Criticism 2: Definition of "Content Relevance"

The second criticism is in response to the assertions made in this dissertation that the content of the items used for syncritic studies must be relevant to the criterion being investigated. There was no standard suggested for determining when item content was relevant to the criterion. At present, I cannot offer such a standard. In fact, however, there are two "guidelines", which were followed in this dissertation, that serve as a partial solution to this problem. In the first place, rational judgment may be used. If an inventory was to be used syncritically as a basis for assisting clients' career explorations, then the appropriate content would be activity, or occupational interests. When this guideline is followed it is assumed that the person making the judgment is well versed about the problem to be studied. The second suggestion is to empirically determine the validity of the items. Although it is clear that chosen item sets must work (or, prove valid) for the intended purpose, a bit of a logical problem is encountered by the empirical solution. The difficulty is that just because a set of items works does not imply
that the content is "relevant" in the rational sense sug-
gested above. Conversely, if an item set doesn't prove
valid, it can't be concluded that the content of the items
is not relevant to the predictive problem. It is hoped that
research will show that content relevance is indeed neces-
sary. However, this end may be difficult to achieve without
a clear standard for judging so-called relevance.

Criticism 3: Matching People Implies that
"It Takes One to Know One"

The third criticism asserts that using syncrisis is
naive since its use is founded on the assumption that, "it
takes one to know one". Seemingly implied in client-thera-
pist matching research, for example, is the notion that ther-
apists are capable only of treating people who are quite like
themselves. There are two primary responses to this crit-
icism. First, syncrisis could still be used even if items
were to be scored for disagreement. There is no technical
difficulty involved in accomplishing this, but it would be
necessary to articulate reasons for so doing. Secondly,
syncrisis results in relative, ordinally arranged orderings
of similarity among people. There is no assumption that two
people must be perfectly matched on their responses to a set
of items before syncrisis may be used for various "matching"
purposes.
Consider the CTM situation. Use of syncrisis merely recognizes that widely different people probably have less basis for developing the sort of mutuality required for effective therapy. There is no assumption that a perfect correlation exists between amount of similarity and counseling outcome. Likewise, no rigid, "it takes one to know one", assumption is implied in the career-matching concept (Kuder, 1977) discussed earlier. Recall that information from many of the people most similar to the client would be dispensed. Clients may then decide for themselves whether information from the various criterion people (which might be personal history, advice to career seekers, etc.) is useful to them. Syncritic scoring of an inventory is not different from standard scoring with respect to this criticism. The primary difference between the two methods is whether the criterion which the client is being compared to is an individual or a group's average response pattern.

Conclusion

This dissertation has perhaps produced more ideas than hard evidence about syncrisis. By this assessment I do not mean that the results gathered in the course of the three studies are themselves unimportant. On the contrary, Experiment 1, for example, produced very important data attesting to the test-retest reliability of syncrisis. But, the yield of the
studies was such that there were numerous opportunities to speculate, opportunities which were not foregone. These speculations were offered as leads to future research projects. The results of future research will determine whether my conjectures have been justified.
REFERENCE NOTES


REFERENCES


Cronbach, L. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.


Krumhansel, C. Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 1978, 85, 445-463.


Whitehorn, J., & Betz, B. A comparison of psychotherapeutic relationships between physicians and schizophrenic patients when insulin is combined with psychotherapy and when insulin is used alone. *American Journal of Psychiatry*, 1957, 113, 901-910.


ACKNOWLEDGMENTS

Members of my doctoral committee were extremely helpful to me throughout the many phases of this dissertation. The criticisms, questions, and ideas of Thomas Andre, Wilbur Layton, Norman A. Scott, and Richard J. Van Iten challenged me to meet high standards. The assistance of Donald G. Zytowski was instrumental to my completing this dissertation. He provided ideas, resources, and data which made this project possible. For three years Chairman Fred H. Borgen has promoted my intellectual development. Without his tutorials I could not have begun to conceptualize and carry out the complex project reported in these pages. Many of the skills I have as a researcher and counseling psychologist are the result of his fine teaching.

Two people not on my doctoral committee also deserve special mention. Fritz G. Kuder originated the idea of career matching which served as the basis for my development of syncrisis. Without his ideas this dissertation would have been on a different topic. Katherine A. Seling produced the term "syncrisis" to denote the procedures used in this dissertation. She also helped with numerous other aspects of the project enabling me to complete it. I was truly fortunate that so many talented people helped me.
APPENDIX A

Questions Asked of Therapists in Experiment 1
Therapist Subject Questions

Below are listed questions concerning your activity interests, personality characteristics, and your therapeutic style. Please give careful thought to each question as it applies to you and prepare a series of answers to the questions which will take about five minutes to deliver orally. Think about these questions for a few days, after which I will contact you. At an appointed time we will go over your responses to the questions and subsequently audio tape you talking about yourself. I am asking you to be frank and complete. Subjects employed in my dissertation study will view your answers, but other persons will not be given access to your answers. Thank you.

1. Explain some of your reasons for entering the field of Counseling Psychology (some might be: Helping others, research possibilities, knowing other psychologists, finding other people interesting, desire to be independent, etc.).

2. What sort of person are you, in easy to understand words? (optimistic?, easy going?, widely social or prefer a few, close friends?, intellectual?, feeling oriented?, warm?, well organized?, liberal - conservative?, confident?, etc. Again, these facets of "personality" are provided only for your convenience, the list is not exhaustive of things you may want to talk about, nor is each characteristic to be
referred to in order, if at all. Please try to be more de­
tailed than a quick, "I'm an optimistic liberal". One way
to accomplish this is to define each of your characteristics
rather than name them. For example, "I'm the type of person
who likes to be around others . . . go to wild parties . . .
gamble . . . .")

3. What kinds of activities do you enjoy? Why? (read­
ing, drag racing, classical music, skindiving, watching TV -
perhaps a favorite program, writing studying, hiking, camping,
sports - football, baseball, whatever, these are, again,
provided only to help you get started formulating your own
answer).

In answering these questions please avoid referring to
"demographic" variables about yourself. Specifically, do not
mention the amount of therapy experience you've had, where you
are from, your age, marital status, income, type of car you
drive, etc. If you wish, you may discuss your orientation to
such things, if you feel it is relevant to an understanding of
you by others. You might, for example, say that a particular
type of previous employment experience which you have had be­
because it was enjoyable or aversive, influenced your decision
to become a Counseling Psychologist. But avoid saying that
five years as an expert witness, therapist, probation officer,
etc. influenced your decision.
APPENDIX B

Rating Sheet and Instructions for Clients in Experiment 1
Subject's Ratings Sheet

Name: ____________________________ Age: ___ Sex: ___ Date: ___

Please adopt the attitude that you are a person with a personal problem seeking psychological help. Examples of personal problems are depression, conflict with another person, worry about career plans, problems with the opposite sex, and so on.

Carefully read the transcripts of the six counselors. They have all provided answers to the same three questions. The purpose in having them answer these questions is to provide you with information about them and what they are like. Using this information respond to the questions below.

1. Please rank the counselors about whom you have read in the order of your preference for seeing them for counseling if you were a person with a problem. It is usually difficult to rank order people - it sometimes seems hard to put someone in last place. But, please keep in mind that even when you put a counselor last doesn't mean there's a huge difference between them and the one you ranked first, although there may be a difference. So please give careful thought to your preferences, and respond according to your feelings. Your list of preferences will be kept confidential.

write the numbers of your choices in the spaces below:

Most Preferred 1. ____
2. ____
3. ____
4. ____
5. ____

Least Preferred 6. ____
2. Circle one of the following numbers to indicate for each counselor how much of a preference you would have for seeing them if you were a person with a problem.

<table>
<thead>
<tr>
<th>Counselor</th>
<th>Much Preferred</th>
<th>No Preference</th>
<th>Do Not Prefer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7  6</td>
<td>5  4  3</td>
<td>2  1</td>
</tr>
<tr>
<td>2</td>
<td>7  6</td>
<td>5  4  3</td>
<td>2  1</td>
</tr>
<tr>
<td>3</td>
<td>7  6</td>
<td>5  4  3</td>
<td>2  1</td>
</tr>
<tr>
<td>4</td>
<td>7  6</td>
<td>5  4  3</td>
<td>2  1</td>
</tr>
<tr>
<td>5</td>
<td>7  6</td>
<td>5  4  3</td>
<td>2  1</td>
</tr>
<tr>
<td>6</td>
<td>7  6</td>
<td>5  4  3</td>
<td>2  1</td>
</tr>
</tbody>
</table>