REDDY, Michael Joseph, 1933-
INTERACTION TESTING IN THE MEASUREMENT OF
MARITAL DISTURBANCE.

The Ohio State University, Ph.D., 1971
Psychology, clinical

University Microfilms, A XEROX Company, Ann Arbor, Michigan
INTERACTION TESTING IN THE MEASUREMENT OF MARITAL DISTURBANCE

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Michael Joseph Reddy, M.A.

* * * * *

The Ohio State University
1971

Approved by

[Signature]
Adviser
Department of Psychology
PLEASE NOTE:

Some pages have indistinct print. Filmed as received.

University Microfilms, A Xerox Education Company
ACKNOWLEDGMENTS

The role of adviser is crucial to any dissertation and my memories of Dr. Francis P. Robinson's guidance are both warm and personal.

There were others whose help at a critical time made maybe more difference than they realized--Don Smeltzer, Ronald Fox and Jaye Nelson. I owe a great debt of gratitude too to all those who went out of their way to secure the cooperation of the control group for this experiment.
VITA

June 14, 1933  Born - Preston, England

1954  . . . .  B.A. (Hons) Classics, Manchester University

1968  . . . .  M.A., The Ohio State University, Columbus, Ohio

1969-1970  . .  Clinical Psychology Intern, University Hospitals, The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Counseling Psychology

Professor Francis P. Robinson

Minor Field: Clinical Psychology

Professors Ronald E. Fox and John Kangas
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgments</td>
<td>11</td>
</tr>
<tr>
<td>Vita</td>
<td>111</td>
</tr>
<tr>
<td>List of Tables</td>
<td>v</td>
</tr>
<tr>
<td><strong>Chapter</strong></td>
<td></td>
</tr>
<tr>
<td>I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>II. Research Background</td>
<td>12</td>
</tr>
<tr>
<td>III. Methodology</td>
<td>42</td>
</tr>
<tr>
<td>IV. Results</td>
<td>53</td>
</tr>
<tr>
<td>V. Summary</td>
<td>91</td>
</tr>
<tr>
<td>Appendix</td>
<td>106</td>
</tr>
<tr>
<td>References</td>
<td>108</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Task efficiency scores of two groups of married couples on the</td>
<td>55</td>
</tr>
<tr>
<td>interactional form of the Wechsler Bellevue</td>
<td></td>
</tr>
<tr>
<td>2. Wechsler Bellevue prorated IQ scores of two groups of married</td>
<td>56</td>
</tr>
<tr>
<td>couples</td>
<td></td>
</tr>
<tr>
<td>3. Split-half reliability coefficients on three subtests of the</td>
<td>57</td>
</tr>
<tr>
<td>Wechsler Bellevue, based on raw scores</td>
<td></td>
</tr>
<tr>
<td>4. Time taken on the interaction form of the Wechsler Bellevue</td>
<td>59</td>
</tr>
<tr>
<td>Comprehension and Similarities subtests by two groups of married</td>
<td></td>
</tr>
<tr>
<td>couples</td>
<td></td>
</tr>
<tr>
<td>5. Mean scores of two groups of married couples on 4 process</td>
<td>61</td>
</tr>
<tr>
<td>variables derived from Wechsler Bellevue interaction protocols</td>
<td></td>
</tr>
<tr>
<td>6. Means, standard deviations and t-ratios for two groups of</td>
<td>62</td>
</tr>
<tr>
<td>married couples on 3 process variables derived from Wechsler</td>
<td></td>
</tr>
<tr>
<td>Bellevue interaction protocols</td>
<td></td>
</tr>
<tr>
<td>7. Mean scores of 50 married couples on process variables derived</td>
<td>65</td>
</tr>
<tr>
<td>from the Wechsler Bellevue interaction test (after Bauman and</td>
<td></td>
</tr>
<tr>
<td>Roman, 1968)</td>
<td></td>
</tr>
<tr>
<td>8. Comparison of group mean scores from 3 groups of married</td>
<td>67</td>
</tr>
<tr>
<td>couples on 5 process variables derived from Wechsler Bellevue</td>
<td></td>
</tr>
<tr>
<td>interaction protocols</td>
<td></td>
</tr>
<tr>
<td>9. Analysis of variance of mean Dominance scores from two groups</td>
<td>73</td>
</tr>
<tr>
<td>of married couples on the Wechsler Bellevue interaction protocol,</td>
<td></td>
</tr>
<tr>
<td>Similarities subtest only</td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>10. Analysis of variance of mean Dominance scores from two groups of married couples on the Wechsler Bellevue interaction protocol, Comprehension subtest only</td>
<td>74</td>
</tr>
<tr>
<td>11. Analysis of variance of mean Dominance scores from two groups of married couples on the Wechsler Bellevue interaction protocol, Similarities and Comprehension subtests combined</td>
<td>75</td>
</tr>
<tr>
<td>12. Analysis of variance of mean Dominance scores of 30 married couples (control group) on the Wechsler Bellevue interaction protocol, Vocabulary subtest only</td>
<td>79</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

In 1961 Jackson et al. described a method of analysis of a family interview which seemed to make it possible to distinguish the communication patterns of the parents of a schizophrenic child from the communication patterns of a normal child's parents, even without the child present. At the same time, however, the same researchers were emphasizing the critical need of better techniques of family diagnosis (Jackson & Satir, 1961). While a variety of techniques exist for the evaluation of the individual family member there were still relatively few such tools for the assessment of the family as a single psychosocial entity.

The following year another member of the same group of family therapists made the following observation (Haley, 1962):

The crucial differences between families would seem to reside in the sorts of transactions which take place between family members; the study of differences becomes a classification of communication patterns in the family. . . . All of the researchers seek some way of describing the unique kind of interactive process observed when these family members are brought together. Further, the
goal is to phrase such a description in a way which would ultimately permit quantitative validation of descriptive statements (pp. 266-267).

In the decade which has followed these beginnings conjoint family therapy has become commonplace, it has become fashionable to speak of the patient as the "identified sick member of the family," and Laing's claim that schizophrenia is a social more than an individual disease has been widely quoted. Ferreira and Winter (1965) bear witness to "the increasing conviction that the family of an individual patient is an abnormal family, different in some ways from a normal family" and that "the investigation of family interaction assumes that this difference is demonstrable and seeks to identify the parameters involved."

The same writers attest further that the last decade has indeed seen a spurt of interest in interaction testing of family members with each other, prompted largely, in their opinion, by the growing popularity of conjoint family therapy (Winter & Ferreira, 1970).

**Interaction Testing**

Interaction testing, as will be seen in Chapter II, can take many forms. Most characteristically it involves the securing of a "consensus protocol," that is, an agreed or joint set of responses from the group being tested as a single entity to a standardized testing instrument. In the spirit of Cattell's (1948) position on the isomorphism of
group and individual processes, it assumes that the scores and profiles from consensus protocols on traditional clinical tests are characteristic of the group (that is, reliable) and predictive of the group's functioning (that is, valid) in ways that are analogous to the scores derived from the same tests when administered to individuals.

In the case of the Bauman and Roman (1967) study, which provides the starting point for the present experiment, the testing format requires, prior to the consensus protocol, the securing of individual protocols from each of the group members. Individual and consensus protocols are then systematically compared. Many traditional or even newly standardized testing instruments can be used with this format. In the Bauman and Roman study the Wechsler-Bellevue intelligence subtests of Comprehension and Similarities were used, as they are in the present experiment. Of all the variations of consensus protocols tried by different researchers none has been more successful in yielding a reliable and meaningful quantitative analysis of the data. A fuller description of this procedure will now be offered.

First of all, the Similarities and Comprehension subtests, from both alternative forms of the Wechsler-Bellevue scale, are administered twice, in the first place to husband and wife separately, and immediately afterwards
to the couple as a group. The protocols are scored and prorated (according to Wechsler's formula for arriving at a full verbal IQ from one or more individual subtests) to give three "IQ"s. The propriety of the use of the term "IQ" may be debated but it was so used in the Bauman and Roman series, and will be used here for the sake of convenience. The three "IQ"s are:

1) A prorated IQ for the husband
2) A prorated IQ for the wife
3) A prorated "interactional" IQ derived from the joint protocol.

A fourth IQ is then derived which might more appropriately have been termed, in view of the literature of group processes, a "synthetic" IQ. Bauman and Roman called it a "potential" IQ. It is derived by summing the scores of the best answers taken from both individual protocols. It represents the optimal use of their joint resources available to the couple.

The final step is to note the difference between the actual interactional score and the potential score which the couple would have made had they used the best answer to each item which one or the other had given in their individual responses. Thus the Potential IQ minus the Interactional IQ yields a "task efficiency" or more accurately a "task inefficiency" score. Roman and Bauman et al.
(1965) found that this score distinguished at a statistically significant level between a group of normal couples and a group of "pathological" couples.

Bauman and Roman also analyze the data qualitatively. Here the interest is on the process rather than the product of the couple's interaction and the mode of operation is not by summary figures but by item analysis. A concrete example will serve to illustrate this phase of data processing. We will take the possible responses of four couples (both separately and as a team) to the question "In what way are peach and plum alike?"

<table>
<thead>
<tr>
<th>Husband</th>
<th>Wife</th>
<th>Interaction</th>
<th>Scored</th>
</tr>
</thead>
<tbody>
<tr>
<td>Couple 1</td>
<td>both fruit</td>
<td>both have skin</td>
<td>both fruit</td>
</tr>
<tr>
<td></td>
<td>and one stone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Couple 2</td>
<td>both fruit</td>
<td>both have skin</td>
<td>both fruit</td>
</tr>
<tr>
<td></td>
<td>and one stone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Couple 3</td>
<td>both fruit</td>
<td>both have skin</td>
<td>both begin</td>
</tr>
<tr>
<td></td>
<td>and one stone</td>
<td></td>
<td>&quot;P&quot;</td>
</tr>
<tr>
<td>Couple 4</td>
<td>both fruit</td>
<td>both fruit</td>
<td>both fruit</td>
</tr>
</tbody>
</table>

The first example is scored "D" for Dominance because one partner's response dominates to the exclusion of the other. It is further noted as D(H) because the dominance in this case is on the part of the husband.

The second example is scored "C" for Combination because elements from both individual protocols were combined in the joint response.
The third example was scored "E" for Emergence because the two spontaneously agreed on the same answer and reinforced that agreement in their joint response.

A further notation is placed on these responses from the interaction protocol. Each can be scored plus (+), minus (-) or zero (0), depending on its value compared with the two individual scores. Again a concrete example will best illustrate this aspect of the scoring. It will be recalled that Wechsler's standard scoring system (which was followed by Bauman and Roman) allows for three levels of response on these subtests. Each answer can be scored 2, 1 or 0 depending on its quality. The responses given on the previous page to the question "In what way are peach and plum alike?" would be scored as follows:

<table>
<thead>
<tr>
<th>Couple</th>
<th>Husband</th>
<th>Wife</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

The interaction response in the first case, as well as being scored D(H) would also be scored plus, because it meets the requirements of a plus score (equal to or better than the best individual score). The second case, as well as being scored C, is also scored plus, for the same reason. The third case was scored E, but is also scored minus because the interaction score is poorer than the better
score. The fourth case was scored R, and is also scored zero because the interaction score is the same as both individual scores. (It might be noted that Reinforcement will always be scored zero since, by definition, it implies that the responses are identical.)

In summarizing this qualitative aspect of the data one may pay attention to the relative proportions of plus and minus answers, to the relative preponderance of one or other process mode or, in the case of Dominance scores, to the relative contribution of husband or wife. Bauman and Roman (1966) found that only two qualitative aspects of the interaction reliably distinguished between the normal and the pathological group, there being significantly more cases of negative emergence in the pathological group and significantly more instances of reinforcement in the normal.

One final aspect of Bauman and Roman's data analysis may be mentioned. They correlated Dominance with various other factors and found that it was overdetermined. In both groups, husbands' responses dominated over wives', the higher IQ partner dominated over the lower, the recorder (the one who actually wrote down the joint answers) dominated over the non-recorder, and in the case of the pathological group the non-patient dominated over the patient.
Purpose of This Paper

The technique evolved by Bauman seems to be both simple and effective. It provides an objective and standardized method of measuring the "product" interaction between husband and wife. It is easily administered and scored and further studies have indicated that it is both reliable and valid. Further, it appears to offer the opportunity for "process" or qualitative analysis of the same interaction.

The merits of this approach to the measurement of interaction will undoubtedly lead to its adoption by other researchers. For the diagnostician it offers a standard series of stimuli enabling him to compare the effectiveness and nature of interactive processes between different couples. As more evidence becomes available some sort of family typology may be able to be constructed on the basis of the numerical scores which interaction testing affords. It can be used directly in therapy as a means of demonstrating certain interpersonal processes through a concrete example a couple can appreciate. At least one clinic to the writer's knowledge uses the Bauman and Roman testing technique as part of initial routine diagnostic practice. The present study seeks to validate this technique and extend its use further. It focuses on two specific problems.
First of all, although Bauman and Roman found significant group differences between the normal and the "pathological" group, how far is this latter "pathological" group typical of disturbed couples? In fact one partner in each pair was currently hospitalized for psychiatric treatment when testing was done by Bauman and Roman. Could one expect to find similar group differences between normal couples and (say) couples who apply for family therapy at an outpatient clinic? Possibly, but maybe not. Yet the Bauman and Roman technique is being used in such settings. Norms are needed for different populations of disturbed couples. Bauman and Roman themselves discovered some slight evidence that where the sick member of their couples was suffering from a psychotic disorder performance on the interaction test was likely to be poorer than the case where the sick member had been diagnosed as neurotic.

It is possible, therefore, that in terms of performance on the interaction test a whole scale of degrees of disturbance or pathology might become manifest. On such a scale where might one expect to place couples who present themselves as a couple at a family outpatient clinic? One would anticipate that the very motivation and implicit admission on the part of both partners that they needed help would likely produce a degree of cooperation and a sense of equality in dealing with interaction testing which
would make their performance hard to distinguish from that of a normal couple. This in fact is the first major hypothesis of this paper. The two populations being compared are a normal population and one composed of couples who applied for joint treatment at an outpatient clinic. It is predicted that there will be a difference in terms of their task efficiency in responding to the Bauman and Roman interaction testing technique. It will also be of interest to determine if the present experimental group will reproduce the scoring patterns of the Bauman and Roman experimental group in terms of their Reinforcement and Negative Emergence scores. One would expect that they would. Again, on the basis of a pilot study for this project, it seemed likely that the clinic population from which the experimental group was drawn would likely score higher on Dominance than the control group. Specific hypotheses will be made to test these expectations.

The second major hypothesis of this paper is concerned only with the qualitative variable of Dominance. Bauman and Roman found that it seemed to be determined by a number of different factors. It is possible, however, that there may be yet another determining factor—the nature of the subtest used. Wechsler's (1958) volume on the measurement of adult intelligence points out that the test as a whole and most of the individual subtests are consistently (though slightly) biased in terms of sex. That is
to say, men typically do better on the scale as a whole and better on five particular subtests, where women do better on three. It is possible that on a subtest which tends to favor one sex the Dominance score for that sex will be greater. Bauman and Roman analyzed their data as a whole. It is possible that analysis of subtest scores separately will reveal an important distinction in Dominance trends.

The two subtests used by Bauman and Roman lend themselves to separate analysis in that the Comprehension test tends to favor men and the Similarities test tends to favor women. This paper will try to determine whether such a bias, built into one particular subtest, will affect Dominance scores. As a further test of this hypothesis an additional subtest was selected, Vocabulary, which usually shows one of the widest sex differentials (in favor of women) of all the subtests, and was administered as a supplementary test to the normal group.
CHAPTER II

RESEARCH BACKGROUND

The study of group processes has been developed at a steadily accelerated pace since the early 1930s. There is no question of even attempting to review here the resulting vast accumulation of literature. It would be useful, however, before considering the more immediate research background to this study, to look briefly at the more remote and broader background of group dynamics, specifically at two areas which have a bearing on the topic of this paper, namely, (1) the question of group versus individual effectiveness, and (2) task-environmental versus social-emotional aspects of group effectiveness.

The Remote Background

The question of which is more effective, the individual or the group (possibly the oldest of all issues in the field of group dynamics) is a less vital one than it was when Lorge et al (1958) reported in an exhaustive summary of studies on the topic that in some cases the product of the best individual will be superior to that of the best group. The question today has been refined. Today, the
question is more likely to be phrased: In what kind of context will which particular kind of group, given which specific task, perform more effectively and/or more quickly than any one member of that group.

Some of the answers to the question have now begun to emerge. Barnlund (1959), for example, points out that the group will be more effective than the individual where an important ingredient of effectiveness lies in the critical resources of the group. In a more thoroughgoing analysis of the problem Collins and Guetzkow (1964) offer a number of propositions relevant to the issue. Two such propositions read:

For tasks involving random error, combining several individual estimates or solutions into a single group product will increase accuracy.

For tasks which involve creating ideas or remembering information, there is a greater probability that one or several persons will produce the information than that a single individual will produce it by himself (p. 54).

This work is highly pertinent to the present study, which involves a comparison of group and individual products. It was essential to provide a task where experience had shown that all other things being equal it was more likely that the group product would exceed that of the best individual product. In the case of the Wechsler-Bellevue Comprehension and Similarities subtests, the pooling of information, solutions and critical resources by two people should almost always result in a higher score
than either one would be capable of alone. If this were not the case it would be impossible to draw any conclusion about the failure of some couples to do so.

A contribution from industrial psychology is relevant in closing this section. Studies such as that of Taylor et al. (1958) have indicated that where productivity or effectiveness is measured along a single dimension and where definite criteria of "correct" answers are available the performance of a synthetic group (one formed artificially by combining the better products of individuals) will be superior to the performance of the interactional group (where the product is sought through consensus of an actual group discussion).

The focal point of the technique adopted in this paper is the interactional IQ, derived from a married couple's consensus answers to an intelligence test. Just as the score of the better individual should provide the lower limit of their joint performance so the potential IQ (synthetically derived by summing the better individual responses to each item) should represent the upper limit. It may be that the precise location of a couple's interactional IQ between these upper and lower limits will prove to be a readily obtained and sufficiently sensitive indicator of degree of disturbance.

At this point we have passed over from a consideration of group versus individual effectiveness to the second
major issue in the literature of group dynamics having a bearing on this paper—the question of the relative impact of social-emotional and task-environmental factors in group effectiveness. What we have been saying in other words is that the Wechsler-Bellevue Comprehension and Similarities subtests keeps the task-environmental factors first at a controlled level and second at a low enough level to be sure that poor performance is much more likely to be due to social-emotional than to task-environmental obstacles.

The social-emotional aspects of group effectiveness are of course complex. The distinction between these aspects and task-environmental aspects arose first, it seems, in the context of leadership when observation of the usually low correlations between "liking" and "task-leadership" in groups led Bales (1950) to posit a double status order in any group, one relating to task leadership and the other to liking, a conclusion supported more recently by Fiedler's (1968) research on the interaction of task and interpersonal factors in group productivity.

Leadership is not a major concern of the present study. It is difficult to speak of power and leadership in a two-member group. It is not, however, entirely irrelevant, especially in view of the Bauman & Roman (1966) findings that Dominance is determined even in the case of a fairly simple and emotionally neutral task such as an
intelligence test by such apparent "status" effects as non-patient vs. patient, recorder vs. non-recorder, male vs. female.

The distinction between task-environmental and social-emotional obstacles in group effectiveness has proved durable too in a wider context than that of leadership alone. In one sense the design of the Roman, Bauman et al. study (1965), as well as of other research to be described shortly, is precisely to present a situation to the group being tested which gives maximum play to interpersonal obstacles to group effectiveness. Consensus requirements in this design may lead to the elimination not only of inferior individual contributions but also of superior. Group members mediate interpersonal rewards to each other which may outweigh the satisfactions of group effectiveness. As Shacter et al. (1960) have pointed out, group cohesiveness does not necessarily enhance group productivity. If the predominant influence of the group is antagonistic to productivity then group cohesiveness will mean a decline in productivity.

Maier (1967) makes the same point when he says that reaching agreement in a group is often confused with finding
the right answer and that it is for this reason that the dimensions of a decision's acceptance and its objective quality must be distinguished. Among the interpersonal obstacles to arriving at the right decision Maier lists the dominance needs of an individual (very relevant to the concerns of this paper), the positive or negative valence of comments made during group discussion and such secondary satisfactions as winning a particular argument.

Collins and Guetzkow (1964) offer another proposition which serves well as a summary statement of this section, though it is phrased in a different terminology:

Groups with different communication nets differed in productivity not because the task is directly more difficult, but because some tasks place heavy demands on the interpersonal relations of the group (p. 67).

The present experiment is so designed that not only is the tasks relatively easy but the demands on interpersonal resources are also relatively light. It is hoped that in such circumstances the instrument used in the experiment will be sensitive enough to register the degree of interpersonal difficulties which would spill over even into such a minimally challenging and apparently non-threatening task.

The Immediate Background

The more immediate research background to this paper lies in the area of the analysis of interaction. For the purpose of clarity three subdivisions of the area may be
distinguished: laboratory-oriented analysis of interaction, interview-centered analysis of interaction and test-centered analysis of interaction.

Laboratory-oriented studies appear to be poorly represented in the area of interaction testing but experiments of Haley (1962) and Reiss (1967) may serve as two examples. In both cases Ss were seated at a round table which was partitioned into booths so they were unable to see each other or to communicate except through written and electronic signals. Haley was investigating the formation of alliances in schizophrenic families, Reiss was investigating concept-formation in schizophrenogenic families.

In the case of interview-centered analysis of interaction three strands can again be distinguished. The oldest and perhaps still the most robust runs from the early work of Robinson (1950) and his associates which dealt almost exclusively with the counselor-client type of interview right through to the more recent work of Truax, Carkhuff and Berenson and others. In this research various dimensions of the counseling interview have been isolated and the attempt made to measure their occurrence. Examples of such dimensions are counselor acceptance of client, responding to the core of the client's remarks, taking of responsibility by counselor and client, degree of leading, clarification, interpretation, approval, urging, amount of client talk-time, depth of client self-exploration.
A second fairly pervasive influence in the area of interview analysis of interaction stems from the work of Bales (1950) who attempted an exhaustive classification of all statements into twelve types divided into the four subcategories of task-oriented questions, task-oriented attempted answers, social emotional positive reactions, and social-emotional negative reactions. Despite reported difficulties with it this scheme has been adopted by many researchers some of whose work will be reported on in more detail below.

The third strand of interview-centered analysis of interaction covers the gamut of idiosyncratic systems of analysis evolved by different researchers with some particular purpose or particular theoretical background in mind. An interesting and apparently successful attempt of this kind is that of Jackson et al. (1961) who divided statements between members of a family into "symmetrical" or "complementary," and disqualifying statements into "sequential" or "incongruent" and in this way was able to identify the greater number of incomplete transactions which he sees as typical of a schizophrenic family.

**Interaction Testing**

The third and final subdivision of work dealing with the analysis of interaction has been reserved for more detailed consideration in a special section because it is
the most closely related to the present study. It has been labeled test-oriented analysis of interaction, or more simply interaction testing because it uses as the unit of analysis the responses of a group to a standardized and often traditional testing instrument. A basic element of the research to be reported here is that it requires a joint or consensus protocol by all the members of the group concerned in response to a standard psychological test. Again, the research reported falls into three areas: projective tests, decision-making tests, and intelligence tests.

1) Projective Tests. The preferred instrument under this heading has been the Rorschach test. Klopfer (1968) believes that this use of the Rorschach as a measure of interaction between members of a group, the "consensus Rorschach" as it is most commonly called, has breathed new life into a perhaps moribund instrument. Credit for this innovative use of an old test is usually given to Blanchard (1959) who from chance observation of an animated discussion among a group of social workers who were examining the Rorschach cards was led to try a similar procedure with the young residents of a detention center. He found that the same boys who were notoriously uncommunicative in traditional individual administration of the test were much more productive in joint discussion of the cards and that he was able to draw from the discussion some meaningful comparisons
with the actual gang behavior of the individuals and some helpful inferences as to the group dynamics involved.

Krauser (1964) notes that "in the last 15 years the focus in understanding personality has broadened from a concern with individual dynamics to an interest in family patterns which play an important part in giving rise to these phenomena." He goes on to report a single case of the administration of a consensus Rorschach (administered before individual inquiry) to illustrate his point. He suggests that the technique could be used as a diagnostic device to distinguish between different types of family. The only quantitative criterion he offers, however, is the number of responses to different cards. His analysis of the protocol, like Blanchard's, is mostly in clinical and interpretative terms.

Kaldegg (1966) likewise reports on a single case. Using the Zulliger alternatives to the Rorschach pictures and the Similarities subtest of the Wechsler Adult Intelligence Scale he was surprised and impressed by the fact that joint testing of the couple concerned gave a completely different prospect for the survival of their relationship than could have been deduced from their individual protocols alone.

Levy and Epstein (1964) are also interested in using the "family" Rorschach (as they call it) as a tool for the construction of a typology of families. The format they use
is to administer the traditional Rorschach separately then leave the couple to discuss their joint protocol, the discussion being tape-recorded. The variables they isolate in assessing results center on the concept of leadership. They are among the first to raise the question of whether or not the joint protocol is in any sense "better" or "worse" than individual protocols. They noted an effect which was common in the joint protocol, an effect in the direction of equilibrium. That is, many of the "sick" responses of the individual protocols were suppressed in consensus.

Another, quite different, approach to the use of consensus Rorschach testing is exemplified in the work of Cutter (1967), with a special emphasis on the prediction of suicide. He notes a case where on two individual responses to one Rorschach card complementary qualities are displayed. Where the mother saw people pulling, the son saw people hanging on. On this basis Cutter points to a system of analysing joint protocols in terms of the content-polarities they evince and the role expectations they imply. In an extended investigation of this principle Cutter and Farberow (1968) administered a consensus Rorschach serially to five groups composed by selecting one alcoholic patient on the basis of convenience and then obtaining a consensus Rorschach involving this individual with (a) three of his friends; (b) three of his roommates; (c) his wife, on two occasions separated by six months and (d) a
high-low status pair. An individual Rorschach was also obtained before and after the series. From this, Cutter evolved a series of profiles of roles and role complements across a series of symbiotic relationships. Analysis is conveniently summarized in terms of the themes evoked by each card and the roles and role complements along with the degree of emergency of each particular role in the consensus Rorschach.

Three other researchers have shown great interest in the consensus Rorschach, Nathene Loveland, Margaret Singer and Lyman Wynne. To some extent they have worked together but their work can be distinguished in part.

Loveland (1963) reports on an analysis of consensus Rorschachs which were preceded by solo administration of the individual test (i.e., a written protocol without an examiner), and followed by traditional administration of the inquiry. In her approach emphasis is given to the manner or process by which two or more persons achieve consensus on a Rorschach card rather than on the number and content of consensual agreements. Her scoring procedure emphasizes three aspects of the transaction: (a) the clarity and vividness of each speaker's communication; (b) his understanding and grasp of the meaning of each speaker's communications, and (c) the affective stand the participants take in relation to each other and to the task.
Loveland (1967) has also been exploring the use of consensus Rorschachs in conventional group therapy (sometimes including the therapist), with individual patient-therapist dyads and with families plus their family therapist. This approach seems promising in evaluating change in communication patterns which might be supposed relevant to the psychotherapeutic process, and in evaluating therapeutic change generally in a situation outside the therapy session itself. Loveland reports on this use of the consensus Rorschach with some 150 differently composed groups ranging from 2 to 30 members.

Singer (1965) used individual parent protocols from Rorschach, TAT, WAIS, Draw-A-Person, Incomplete Sentences Blank testing and other data to predict with dramatic success the existence of psychiatric disorders in their children. She also used the consensus Rorschach but found that surprisingly she gained more knowledge of parents' communication patterns in the standardized individual form of the test (where the examiner acted apparently as a totally passive recipient of typical communications) than she did from the joint protocol, where some pertinent information seemed to have been lost. In her assessment of Rorschach responses, Singer attends to what she calls closure problems, disruptive behavior, and peculiar verbalizations. These she further relates to patterns of handling attention and meaning which would impair a growing
child's capacity for selective attention, purposive behavior and subjectively meaningful experiences.

Wynne (1968) as well as participating in some of the research of both Loveland and Singer has also offered a useful overview of some of the work with the consensus Rorschach. As the reader will have been aware there is a serious problem with the variety of modes of administration, of scoring and of theoretical orientation in interpreting consensus Rorschachs. Wynne underscores some of these problems: the order of administration of individual and consensus Rorschach, the degree of participation by the tester, the number of cards used, the number of responses expected from each card, the method of judging when consensus has been reached, etc.

From the point of view of this paper it is disappointing that so few of the studies reported have systematically examined or at least reported the differences between individual and consensus Rorschachs. Two of those who have so reported (Singer, 1965; Levy and Epstein, 1964) seem to imply some loss of material in the consensus Rorschach—a phenomenon which is not, however, without its own diagnostic implications.

One more entry needs to be made under the heading of projective tests. Winter and Ferreira (1965, 1966, 1967, 1970) have reported on an interaction technique using the TAT. In this format 3 cards chosen from the TAT series are
presented to three members of a family. Their task is to invent a story which links all three cards. They take turns in reporting the consensus story to the examiner. The data has been examined in three different ways, first by means of Arnold's sequential story analysis (adapted) which distinguished well enough between the normal group and the disturbed group but not at all between the subclasses of the disturbed group (schizophrenic, maladjusted and delinquent). Secondly the data were analysed by means of Bales' (1950) Interpersonal Analysis system, and again the analysis added little to the already diagnosed group differences. More recently, using a wholly different approach, Winter had his raters assess such factors as the number of seconds spoken by each member of the family and the number of seconds of overlap and of silence, and made his analysis in terms of: (1) time taken to reach a decision, (2) time taken by different family members, (3) the sequence of comments (who speaks after whom), (4) the amount of overlap talking, and (5) the amount of silence. Results again were inconclusive.

2) Decision-making tests. The basic formula employed in studies of this kind is the presentation of a conflict situation and the requirement that the group come to an agreed solution. The studies of Strodtbeck (1951), Kenkel (1959, 1961), Titchener et al. (1963), and Mishler and Waxler (1966) all follow this model. A problem, usually
with some domestic bearing (such as the expected dating behavior of teenage children, the spending of an unexpected $300 bonus), or a series of such problems is given to the members of a family individually. They are then made aware of the differences of opinion in their answers and asked to resolve the differences. The discussion is tape-recorded and analysis is made of this interaction. Beyond this, unfortunately, the studies quoted are not comparable. Mishler and Waxler are interested in relating certain patterns of interaction to schizophrenia and they use a great variety of analytical methods; Titchener is interested in relating the family process to the formal personality characteristics of the individuals concerned and expresses and uses as units of interaction 'formal' variables such as 'communication' or 'control,' and 'content' variables such as 'value orientation' or 'anxiety'; Kenkel, finally, uses the Bales (1950) method of interpersonal analysis and tries, with little success, to correlate these categories of communication with various personality traits--Dominance, Persistence, Self-confidence, Conservatism, Authoritarianism, and so on.

A slightly different approach was used by Goodrich (1963). He had married couples sit opposite each other, though hidden from each other's view by a panel on each side of which were 30 smaller panels of different colors and shades, all numbered. The couples were then
simultaneously presented with a colored card and asked to match it with one of the colors in front of them. The task in itself was not difficult but in the case of 10 of the 20 matching problems the numbers of the colors did not correspond from one master-panel to the other, making disagreement inevitable. Goodrich analysed the couples' discussions in terms of such descriptive variables as 'tolerance for ambiguity,' 'anxiety,' 'confusion' and different ways of handling or avoiding conflict. He uses only one quantitative measure—the number of agreed solutions where agreement was in fact an impossible demand. The theoretical base of Goodrich's research was Erickson's theory of life stages. Ss were all between the 4th month after their marriage and the 3rd month after the birth of their first child. The aim was to show that mastery of the intimacy stage was predictive of mastery of the generative stage.

In a totally different approach to analysis of the same data Ryder and Goodrich (1966) extracted 17 variables—such as number of statements made by couple, number of times discussion initiated by one partner, amount of laughing, number of disapproving remarks, etc. From a factor analysis of these variables three major factors were extracted which the authors thought might allow for a diagnostic classification of couple type. The three variables were (1) an affectivity-rationality dimension;
(2) length of disagreement, and (3) tendency to eliminate
disagreement by distortion and error.

The most persistent researchers in the area of
decision-making tests are Ferreira and Winter (1963, 1965,
series of experiments which were dubbed initially by
Strodtbeck (1951) the technique of "revealed differences"
Ferreira and Winter call their technique one of "unre-
vealed differences." The series began in the hope that
"out of such contrasts (in decision-making processes) the
process of family decision-making would permit us a way
of talking about distinct family groups, and a first tool
towards the conceptualization of transactions and
relationships.

The format is extremely simple. It first appeared
in embryonic form as a group of 16 situations where each
member of the family had to rank-order three items (for
example, preference for foods in a restaurant). After writ-
ing down their individual choices the members of the family
were invited to sit down together and (without revealing
their own original choices) come to some common agreement
about the sixteen situations. Decisions were character-
ized as (1) unanimous, (2) majority, (3) dictatorial, or
(4) chaotic.

In the final form of the Ferreira and Winter system
there were only seven situations, but there were ten
alternatives for each and Ss were asked to rank order their preferred three and their least preferred three. Out of 17 hypotheses concerning differences between normal and disturbed families, three variables were isolated which seemed to distinguish reliably between the two populations—spontaneous agreement (SA), choice fulfillment (CF) and decision time (DT). "The stability of these measurements indicates that the study of the family decision-making process may well become a Promised Land to the investigator of family interaction inasmuch as there seem to be—in the variables of Spontaneous Agreement, Decision Time and Choice Fulfilment—satisfactorily stable indicators of meaningful coordinates of family functioning" (Ferreira and Winter, 1966, p. 355).

Later studies included not just the original triad of mother, father and child, but a second child; and any one member of the family or even the family as a whole might be the designated "sick" one. Again there was more choice fulfillment and more spontaneous agreement in the normal group and a longer decision time in the disturbed. More recently a second form of analysis of data has been added. The difference in CF and DT scores was traced to the fact that in normal family discussion more valid and explicit information is exchanged, and there is less time wasted in silence. Most recently Ferreira and Winter have been pooling the twin series of studies they have piloted
in an effort to find if the variables they have isolated as distinguishing the normal from the abnormal interaction overlap in some way or could be reduced to simpler factors (Winter and Ferreira, 1970). A similar study was made by Singer and Robey (1967).

3) **Intelligence tests.** This final section deals exclusively with the Bauman and Roman experiments which give rise to the present study. They will be briefly recapitulated below but it will be useful at this point to find some way of summarily categorizing the body of research just described, a way which will conveniently situate the Bauman and Roman approach along the various dimensions which are represented by this previous research. Five such strands have been selected which permit a systematic comparison of the points at which the Bauman and Roman series either run parallel to or diverge from the foregoing studies.

(a) Attention has already been drawn to the disparate instruments, methodology and theoretical underpinnings of three different studies. It should also be noted that they have widely differing aims. Some are simply correlational studies relating performance on a consensus protocol to other personality variables (Kenkel, 1959, 1961); others are seeking the dimensions of a facilitative relationship particularly that of therapist to client (Robinson, 1950); some have tried to use the consensus
protocol as an index of therapeutic change (Loveland, 1967); many would like to identify the parameters of a family typology (Levy and Epstein, 1964); others again have sought to bring the results of interaction testing directly into therapy. Richman, for example, cites the case of a couple whose individual Rorschachs showed a marked preference in one case for C (color) responses, and in the other for M (movement and fantasy) responses, but whose joint protocol showed an absence of both kinds of response and a preference for the F (more strictly rational) kind of response. He used this as a demonstration to the couple of the way in which their individual richness was suppressed and impoverished in their joint communication and action (Bauman and Roman, 1968). The bulk of the work quoted above has, however, been the work of clinicians intent on supplementing and sharpening their tools of diagnosis as an adjunct to therapy. Almost all of the Rorschach research has been of this kind and the Bauman and Roman studies must likewise be placed in this category.

(b) In terms of Ferreira and Winter's (1968) distinction between the techniques of revealed versus unrevealed differences, the Bauman and Roman approach is that of unrevealed differences—that is, the individual members of the family do not reveal nor are they told of the differences in their individual protocols before they are invited to agree on their joint protocol.
(c) Almost all of the research already quoted could be subsumed under the heading of "decision-making," and in a sense all the Ss are required to do is to come to some agreement. A distinction sometimes has to be made, however, between decision-making and "problem-solving," as suggested by Costello and Zalkind (1963). Although the task set by the Bauman and Roman technique does not entirely match Costello and Zalkind's definition of problem-solving, it must be distinguished from the other (decision-making) techniques because it comes closer to requiring the couple to come not just to some conclusion but to some particular (that is, correct) solution.

(d) This section and the following mark the most important comparisons between Bauman and Roman's experiments and other research. In the latter, three different emphases can be discerned. A great part of this previous research focuses almost uniquely on the tape-recorded interaction of the family or group being tested as they try to arrive at a consensus. This approach might have been included simply under the heading of interaction analysis, except that it makes use of a standard situation or psychological test as the stimulus for interaction. The second emphasis is in fact on the consensus protocol itself (which was virtually ignored in the first approach). Blanchard's (1959), Loveland's (1963) and Cutter's (1967) work are all good examples of this approach. The third
emphasis involves some use of both the individual and the consensus protocols. Levy and Epstein (1964) and Singer (1968) both compare the relative diagnostic value of individual and consensus protocols. Ferreira and Winter (1963, 1965, 1966, 1968) use the individual protocols consistently as a reference point for their assessment of the consensus protocol. In this they approach most closely the Bauman and Roman technique which alone uses a systematic comparison of individual and consensus protocols as the unique focus of assessment.

(e) In all the research quoted three levels of analysis of interaction can be distinguished. The first level is a level of formal or stylistic analysis in terms of such factors as number of times discussion initiated by one or other partner (Ryder and Goodrich, 1966), number of seconds of talk by any one group member (Winter and Ferreira, 1970), etc. The second level of analysis would be represented by designation of a statement by such terms as "giving information" (Bales, 1950) or "clarity of the speaker's communication" (Loveland, 1963). This level seeks to label a type of communication. A good part of the research quoted uses this level of analysis at some point, and so do Bauman and Roman in part. The third level of analysis is more in terms of implied personality dynamics. This is little used but one example is that of Titchner et al. (1963), a part of whose analysis is in terms of
"value orientation," "anxiety," etc., implied by a statement.

Just how these different levels might be named is not important here. Some would draw distinctions between process and content analysis, others between formal and content analysis, Blau and Scott (1962) would probably categorize the first level as "interaction" and the second and third as "communication." What is important here, merely for the purposes of clarification of the uniqueness of the Bauman and Roman approach, is that all three levels can be subsumed under the heading of "process" analysis in contrast to Bauman and Roman's "product" analysis. This requires a word of explanation.

Vidich (1956) pointed out that "the most efficient way to stimulate an episode of interaction is to present a problem situation requiring a solution and final agreement." All of the interaction testing described so far has attempted to provide just such a problem-stimulus situation which is standardized and thus allows for systematic comparison of responses between individuals, families or groups. The unique element in the Bauman and Roman approach has already been indicated, namely that there is only a small array of correct responses to the stimulus problem. Further, these few possible solutions can be assigned a definite numerical value--by means of Wechsler's scoring system which assigns a score of 2 points, 1 point or zero
points to each answer. Thus the Bauman and Roman technique allows not only for qualitative variables which can be assessed from the content of the consensus protocol (the second level of process analysis described above) but also for a quantitative assessment of the product represented by the couple's joint protocol. This system appears to be unique, though as long ago as 1964 Kaldegg was expressing surprise that the possibilities of this particular method had not been more extensively investigated. In none of the research cited, however, was any system devised for rating the consensus protocol as in any way "better" or "worse" than an individual protocol.

Ironically enough the only researchers who have attempted to devise such a system of rating other than for intelligence tests are Bauman and Roman themselves. They began their research using a variety of psychological tests, such as the Wechsler-Bellevue, the Rorschach, the TAT, Figure Drawing, the Szondi test, etc. They compared consensus Rorschachs, for example, taken at 9-month intervals from group therapy patients, and had a clinician rate the later one in terms of improved group psychosexual adjustment (using the Harrower Multiple Choice Rorschach). In a further pilot project \( (N = 2) \) they used the Davidson (1950) Adjustment Scale for rating Rorschach psychograms, which allows for a comparison in terms of "better" or "worse" adjustment between different protocols (Bauman and
Roman, 1960). At this point they were approximating the method which they have used since with their preferred instrument, the Wechsler-Bellevue intelligence test. A brief overview of their work with this test will complete this chapter.

The format of the Bauman and Roman approach to interaction testing involves first of all the standard administration of the Similarities and Comprehension subtests of the Wechsler-Bellevue intelligence scale. Both forms are administered, giving the equivalent of a total of four subtests. The raw scores are scaled and prorated according to Wechsler's formula to give a Verbal IQ. Immediately after individual testing, the couple are seated at a table across from the examiner, who places an answer booklet and a pencil between them and asks them to complete an identical protocol, this time completing it by way of consensus on each item. The interaction or consensus protocol is scored in an identical manner to the individual tests, and an Interaction IQ is thus derived. A further score is then derived by summing the scores on each item, using the best answer in each case from one or other of the individual protocols. Theoretically this is the best the couple might have done had they made maximum use of their resources. Scores are again scaled and prorated in identical fashion and a Potential IQ is computed. A unitary score for each couple is then obtained by subtracting their actual score
in interaction from the score representing their maximum possible score, by subtracting, that is, the Interaction IQ from the Potential IQ. This score is called the "task efficiency" score, and is the score used in assessing the quantitative aspect of the joint product.

Qualitative scores are then derived, or process scores (as opposed to product scores). Each item from the interaction protocol can be scored in two ways. First it can be assigned a quality of plus, minus or zero, depending on whether it equalled or surpassed (plus) the better individual score, was inferior to the better individual score (minus), or was equal to both scores (zero).

Each item from the interaction protocol is then scored for Dominance, if one partner's individual response dominated to the exclusion of the other's; for Combination, if elements of both individual responses were used; for Emergence, if a wholly new idea appeared; and for Reinforcement, if the response was the same on all three protocols (both individual protocols and the consensus).

Fifteen possible categories are thus available for classifying responses in this way—Husband Dominance, Wife Dominance, Combination, Emergence and Reinforcement, each of which can be subdivided into plus, minus or zero cells.

In their initial comparison between 50 normal couples and 50 couples each including one spouse currently hospitalized for psychiatric reasons Roman, Bauman et al. (1965,
1966) found that the task efficiency score and the categories of Reinforcement and Negative Emergence distinguished significantly between the two groups. Split half reliability measures indicated that the consensus or interaction IQ had roughly the same order of reliability as individual IQs ($r = .85$), and that the qualitative variables (Dominance, etc.) showed an order of reliability ranging from .51 to .84.

The next year Bauman and Roman published an administration and scoring manual for their test procedures (Roman and Bauman, 1964). They also analyzed more extensively the data on Dominance alone and found that it appeared to be overdetermined. Husbands responses dominated over wives, the higher IQ partner over the lower, the recorder (the one who used the pencil for the joint protocol) over the non-recorder, and the non-patient over the patient (Bauman and Roman, 1966).

A follow-up study was made two years later (Roman and Bauman, 1966) using 16 randomly selected couples from the original group of 50 "pathological" couples. They found that while there were no significant changes in individual IQs, there were significant increases in the Interaction IQ (a mean of 10 IQ points)—a remarkable finding in that it appears to reflect changes in the couple as a unit which are not reflected in nor deducible from individual IQ data. Bauman and Roman note their impression
that this reflects changes in the couples' capacity to constructively handle their differences.

The same study showed further that there were a significant increase in Reinforcement scores and a significant decline in Negative Emergence scores—the two factors which they had previously found distinguished normal and abnormal couples. Equally notable was the fact that the expatient's Dominance scores (significantly lower than partner's at the time of original testing) had in the interim increased significantly. One last indication from the data was noted. It appeared that the "task efficiency" score of couples with a psychotic partner were significantly lower, in both the original and the follow-up studies, than couples with a neurotic partner.

It is this latter finding that has prompted the present paper, and it will be useful to recall in conclusion the specific intent of this study. Bauman and Roman used as their designated "pathological" couples married partners one of whom was currently hospitalized in a psychiatric setting. It might be supposed that interaction between such couples would be seriously hampered and the Bauman and Roman studies seem to show that such difficulties could be accurately registered by their interaction testing technique. It might, however, be supposed that couples of whom neither was currently designated as psychiatrically sick could indeed have difficulties in
interaction but not of the same magnitude (in terms of the "task efficiency" score) nor of the same kind (in terms of qualitative variables such as Dominance, etc.) What of couples, for example, who applied for joint marital counseling in an outpatient setting? Would the Wechsler-Bellevue tests prove sensitive enough to register degrees of difficulty in interaction among couples such as this—at least sufficiently to distinguish their performance from that of normal couples? It is to answer this question that the present study was chiefly designed.
CHAPTER III

METHODOLOGY

The purpose of this study is to compare performance of two groups (a group of couples sufficiently disturbed to apply for marriage counseling and a control group of normal couples) on the interactional form of the Wechsler Bellevue Comprehension and Similarities subtests.

**Subjects.** The experimental or maritally disturbed group was composed of couples who applied for joint counseling on an outpatient basis at the Ohio State University Family Service Unit in the Department of Psychiatry (N = 33). In terms of degree of emotional disturbance they were thus supposed to lie somewhere between the general population and the Bauman and Roman (1967) experimental group where one partner of each couple was receiving in-patient psychiatric treatment. The experimental group was tested as soon as they applied for counseling. This included MMPI testing as potentially a guide to estimating the degree of emotional disturbance currently experienced by these couples. Sometimes other members of the family such as a child or children were presented and tested along with the husband and wife. In such cases, of course, the data
could not be included in the present study which was concerned uniquely with the interactional performance of husband/wife dyads. Three other couples whose data was incomplete was also excluded from this experiment.

The so-called normal group was recruited through personal contact ($N = 30$). To be more specific couples were recruited through personal contact on the part of mutual acquaintances of $E$, since it was desired to exclude couples known personally to $E$. Testing was topped when the cooperation of 30 couples was secured. The term "normal" was operationally defined as designating couples who were (a) not presently receiving conjoint therapy; (b) not anticipating the need for such therapy; and (c) not under individual psychiatric treatment. As a further check on the "normalcy" of this population the MMPI was administered to them. Deviations from the normal range of profile were rare enough and minor enough to confirm that this was in fact a representative group of individuals from the general population.

The most important element of selection of this normal sample after its definition was to secure a population which was roughly matched with the disturbed group (whose characteristics were a given and had to form the basis of comparison). It would have been a virtual impossibility to match the groups on a one-to-one basis on every variable which might possible influence marital
interaction. At the other end of the scale a random sample of the normal population would not necessarily have offered the best control group. There is nothing to impose the conclusion that the sample of couples who applied for counseling at the clinic were themselves a cross-section of the normal population on every other dimension beyond the fact that they had applied for help. It was more important, therefore, that the control group match the experimental group in a rough way and that it should not include a preponderance of any couple type which might bias the sample in some way. It should not, for example, include a large proportion of couples of a certain age, of couples with a large family or no family at all, of couples of one particular socioeconomic level, of couples with some particular religious affiliation, of couples who had been married only a short time or who had been married for many years. It was felt that such matching was in fact achieved. Apart from the fact that no religious group was over-represented in either group, the average age of the normal group was 38 and 33 for the disturbed. Probably by the same token the normal group had a similar slight edge in terms of average number of years married (15 compared with 11) and of children in the family (average 2.8 compared with 2.1). Most importantly, using Hollingshead's (1957) system based on the twin criteria of education and
occupation which assigns a numerical score to each couple, the normal group showed a socioeconomic level score of 27.20 and the disturbed group an average score of 29.06.

Procedure. It appears that in the Bauman and Roman studies testing of the normal group was conducted through the mail. In the present case all Ss were tested personally by E or one of two assistants, who were present throughout testing but took no part whatever in the completion of the protocols. All testing for both groups took place in the same room at the clinic, with the exception of a small number of normal couples (N = 8) who could only be tested in their own homes.

Testing procedures. First of all, the couples were asked to fill out a single biographical data sheet routinely administered to all patients at the clinic (see Appendix A).

They were then given the test booklet proper, one copy to each partner. This booklet, prepared specially for this study, contained the Wechsler-Bellevue Comprehension and Similarities subtests (Form I and Form II). Questions from these two forms were interleaved. That is to say, Question 1 (a) was question 1 from the Comprehension subtest, Form I, Question 1 (b) was question 1 of the Comprehension subtest, Form II, Question 2 (a) was question 2 of the Comprehension subtest, Form I, and so on. The total number of items in this combined form is 44. In the case of the normal couples an extra sheet was included at the end
containing the 37 items of the Wechsler Adult Intelligence Scale Vocabulary subtest. Couples were simply asked to complete the protocol as well as they could and not to consult with each other. No mention was made of time restrictions and none were imposed.

When both partners had indicated they were finished they were allowed a two-minute break. They were then asked to sit together, another blank protocol (identical to the individual protocols) along with a pencil was placed between them. They were given the following instructions:

"Each of you took this test alone. The purpose now is to determine whether both of you can do better together than each was able to do alone. Read each item and both of you are to discuss the question and agree upon the best answer."

E gave no help in making the choice of recorder. As a matter of fact the matter was settled by all couples in both groups in a matter of seconds with no apparent difficulty, though in a few cases the pencil was transferred later from one to the other, again without any indication of this causing a problem. E made a note of recorder and also (on the basis of Ferreira and Winter's 1965 research) timed this part of the testing procedure.

At the conclusion of the testing session the couple were handed an MMPI booklet along with two answer sheets. A brief explanation of the use of these materials was
given along with the admonition that they were to be completed without consultation, and the request that they be returned within a week in the stamped, addressed envelope included.

Couples were guaranteed complete confidentiality with regard to all test, biographical data, etc. They were requested not to put their name on any of the protocols or the biographical data sheet which were serially numbered for scoring purposes. Couples were also promised a one-page report summarizing results of the experiment when this was completed. They were also given the opportunity to inquire about their own individual results if they so wished. At testing time they were invited to make a note of their serial number so that they could claim the appropriate scores.

Testing procedures were identical for both groups.

Scoring. Individual protocols were scored according to the criteria laid down by Wechsler (1944, 1946). That is, each item is assigned a score of 2, 1 or 0 according to its quality. Scores for individual items are then totalled and scaled—separately for each subtest and each Form of the test. The two scale scores for each Form, in two separate operations, are averaged and multiplied by 5, with the corrections required by Wechsler for Form 2 made independently. The totals of these two operations are then themselves averaged; to transform this final prorated
weighted score to a verbal IQ reference was made to the proper intelligent quotient in Wechsler's manual according to the subject's age.

The Interaction protocol is scored in an identical manner as though it were the product of a single individual.

A fourth protocol is then artificially synthesized by taking only the score for each item which represents the best effort by either partner. This protocol is scored again in identical fashion to the individual protocols. In the case of both the interaction and the synthetic protocols the "subject's age" is determined by the age of the older partner (the one which will yield the higher IQ).

The last stage of scoring elicits a single unitary score for each couple, which is called the "task efficiency" score. This is computed by subtracting the interaction IQ from the synthetic (or potential) IQ.

Scoring was done by E. Correlations with two other experienced raters showed figures of .840 and .952, respectively (10 cases, Spearman rho).

In accordance with the criteria laid down by Bauman and Roman in their administration and scoring manual (1964) the interaction protocol for each couple was also scored for the qualitative variables of Dominance (twice, once for each partner), Combination, Emergency and Reinforcement. In this process each item is scored in two
different ways. First it is assigned to one or more of the following cells, according to these criteria:

An item is scored for **Dominance** if it contains the original response of one partner, to the exclusion of the other's response.

An item is scored for **Emergence** if it includes an idea which was not represented in either of the individual's original responses.

An item is scored for **Combination** if it includes elements of both partners' original responses.

An item is scored for **Reinforcement** if it reiterates a response which both partners had originally given.

Note that in the case of the Dominance response the item is assigned either to the Dominance (Husband) cell or the Dominance (Wife) cell depending on the author of the dominant response.

Each item is then assigned a value of plus, minus or zero. It is denoted plus when its score (according again to Wechsler's criteria) is equal to or surpasses the better individual response. It is scored minus when it is inferior to the better individual response. It is scored zero when it is the same as both original responses. By definition, therefore, Reinforcement can only be scored as zero.
Scores for each couple's interaction protocol are then summed and can be conveniently laid out as follows (a hypothetical case):

<table>
<thead>
<tr>
<th></th>
<th>Plus</th>
<th>Minus</th>
<th>Zero</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominance (Husb)</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Dominance (Wife)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Combination</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Emergence</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Reinforcement</td>
<td></td>
<td></td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>4</td>
<td>30</td>
<td>44</td>
</tr>
</tbody>
</table>

The last phase of scoring involves only the Dominance scores. Here the cell totals are simply broken down by subtests, where in the above totals they had been combined. In the case of the normal group, the Vocabulary scores are also totalled separately. In no case are these Vocabulary scores ever combined with other subtest scores.

Hypotheses. The hypotheses fall naturally into three areas, one concerned with the main product or quantitative analyses of the data, one related to the principal process or qualitative variables investigated, and the third concerned with a more extensive analysis of the single Dominance variable. The present group of maritally
disturbed couples might be considered less emotionally disturbed than the Bauman and Roman group (where one spouse was hospitalized for psychiatric care). At the same time they might be expected to show some differences from the normal population, lying somewhere between the extremes of acutely disturbed and normal. The hypotheses, however, are cast in the null form (no differences expected) for statistical testing.

Hypothesis 1:1. There will be no difference between groups with regard to "task efficiency" score.

This will be tested for statistical significance with the t-test for difference between group means (Hays, pp. 319-322).

Hypothesis 1:2. There will be no difference between groups with regard to total time taken to complete the interaction protocol.

This will be tested for statistical significance with the t-test for difference between group means.

Hypothesis 1:3. There will be no relationship between task efficiency scores and time taken on the interaction protocol, for either group.

This will be tested for statistical significance by means of the Pearson Product-moment Coefficient (Clarke, Coladarci and Caffrey, pp. 110-119).
Hypothesis 2:1. There will be no difference between groups with regard to mean Dominance scores of each.

Hypothesis 2:2. There will be no difference between groups with regard to mean Reinforcement scores of each.

Hypothesis 2:3. There will be no difference between groups with regard to mean Negative (or Minus) Emergence scores of each.

All three parts of Hypothesis 2 will be tested for statistical significance with the t-test for difference between group means.

Hypothesis 3. Dominance scores will not be determined by sex, by intelligence, by group differences in degree of marital discord, nor by the nature of the specific Wechsler Bellevue subtest used.

This will be tested for statistical significance by means of a 3-way analysis of variance performed on A) Dominance scores of each subtest separately; B) total Dominance scores from both subtests combined.
CHAPTER IV

RESULTS

The first main thrust of this investigation was to ask if it was legitimate to regard couples who come to an outpatient clinic for conjoint therapy as representing a different population from a "normal" population, in terms of the couples' performance on a measure of emotional adjustment and on a measure of task efficiency, namely, the MMPI and the consensus Wechsler Bellevue respectively.

Results of MMPI Testing

The MMPI did in fact distinguish the two groups fairly sharply. It was administered to the normal group in the first place to ensure that the control group did indeed represent the "normal" population. It was administered to the maritally disturbed group on the supposition that they too represented a cross-section of the population, a cross-section which was experiencing difficulties in marriage but which need not be expected to manifest, except perhaps occasionally, any personal emotional pathology. This supposition was not borne out in the event.
A certain caution is necessary in interpreting the results of MMPI testing. It is not an instrument which lends itself to a facile analysis of degree of disturbance. Even the normal range cut-off points on the clinical scales allow for exceptions in the case of some subgroups of the normal population; and some scales are more sensitive to transient moods than others. Single small elevations on such scales must be regarded as offering only the most tentative indications of pathology—a "small" elevation being defined, for the purposes of this discussion, as lying between a T-score of 70 (the usual cut-off point from the normal range) and a T-score of 75.

In the present experiment the normal group (with 30 couples) showed MMPI results as follows. Three couples did not return their protocols. Of the remaining 27, 19 couples produced profiles where both were comfortably within the normal range, 6 showed profiles where one partner only showed a small elevation and on only one scale, 1 showed profiles where both partners had small elevations on a single scale. Only two people out of the normal group had an MMPI profile which clinically would suggest serious or acute pathology (in terms of multiple and considerable elevations on the clinical scales). As it so happened, they were husband and wife.

In terms of the rough kind of comparison we are instituting between the two groups, the maritally disturbed
group showed almost the contrary of the trends indicated by the normal group. In their case, out of 66 individuals no less than 30 showed signs of acute emotional disturbance (again in terms of considerable or multiple elevations on the clinical scales) compared to the 2 in the normal group. And only 5 couples showed no sign whatever of emotional disturbance, compared to 19 couples in the normal group. There is no question of a statistical comparison between the two groups, but taken as a whole and in terms of a clinical interpretation of their respective results they must be regarded generally as representing two different populations.

Results of Interaction Testing

This difference in the composition of the two groups was further tested through the procedures which have been described at some length in this paper, namely the interactional form of the Wechsler Bellevue intelligence scale. Table 1 sets out the relevant scores.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>$\bar{X}$</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>30</td>
<td>6.533</td>
<td>5.124</td>
</tr>
<tr>
<td>Maritally disturbed</td>
<td>33</td>
<td>6.787</td>
<td>4.115</td>
</tr>
</tbody>
</table>

*Derived by the formula of Bauman and Roman (1967) and based on IQ prorated, according to Wechsler's (1944) procedures, from Comprehension and Similarities subtest scores.
On the basis of the data set out in Table 1, Hypothesis 1:1—that there would not be a statistically significant difference between the normal group and the maritally disturbed group in terms of "task efficiency score"—could not be rejected. Statistically, the difference between the two groups is minimal (t = 0.215; p > .40), with the normal group performing only slightly better than the disturbed couples.

The reliability of the scores on which the "task efficiency" scores were based is set out in Tables 2 and 3.

TABLE 2.—Wechsler Bellevue prorated IQ scores of two groups of married couples.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Form I</th>
<th>Form II</th>
<th>Pearson r: Form I vs Form II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>X</td>
<td>S.D.</td>
</tr>
<tr>
<td>a) normal group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Husband IQ</td>
<td>30</td>
<td>109</td>
<td>8.78</td>
</tr>
<tr>
<td>Wife IQ</td>
<td>30</td>
<td>110</td>
<td>9.46</td>
</tr>
<tr>
<td>Interaction IQ</td>
<td>30</td>
<td>114</td>
<td>8.06</td>
</tr>
<tr>
<td>Potential IQ</td>
<td>30</td>
<td>122</td>
<td>8.10</td>
</tr>
<tr>
<td>b) disturbed group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Husband IQ</td>
<td>33</td>
<td>107</td>
<td>12.09</td>
</tr>
<tr>
<td>Wife IQ</td>
<td>33</td>
<td>108</td>
<td>16.37</td>
</tr>
<tr>
<td>Interaction IQ</td>
<td>33</td>
<td>116</td>
<td>9.86</td>
</tr>
<tr>
<td>Potential IQ</td>
<td>33</td>
<td>124</td>
<td>10.15</td>
</tr>
</tbody>
</table>

*aCorrected by Spearman-Brown formula.
TABLE 3.—Split-half reliability coefficients on three subtests of the Wechsler Bellevue, based on raw scores.

<table>
<thead>
<tr>
<th></th>
<th>Normal Group</th>
<th></th>
<th></th>
<th>Disturbed Group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comprehension</td>
<td>Similarities</td>
<td>Vocabulary</td>
<td>Comprehension</td>
<td>Similarities</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>Husband</td>
<td>.535</td>
<td>.695</td>
<td>.952</td>
<td>.503</td>
<td>.770</td>
<td></td>
</tr>
<tr>
<td>Wife</td>
<td>.571</td>
<td>.498</td>
<td>.925</td>
<td>.724</td>
<td>.899</td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>.632</td>
<td>.616</td>
<td>.838</td>
<td>.569</td>
<td>.851</td>
<td></td>
</tr>
<tr>
<td>Potential</td>
<td>.468</td>
<td>.686</td>
<td>.921</td>
<td>.613</td>
<td>.871</td>
<td></td>
</tr>
</tbody>
</table>

aCorrected by Spearman-Brown formula.

There is no immediately obvious reason why these reliabilities should be lower than those reported by Bauman and Roman (1968) which ranged between .82 and .88. Part of the difference undoubtedly lies in the slightly lower means and relatively higher standard deviations reported by Bauman and Roman which would tend to boost reliability coefficients to some extent. Slight differences of interpretation of scoring criteria and chance factors may also have played a part. In any case the lack of predicted differences in "task efficiency" scores is not necessarily due to low reliabilities. It should be recalled that the computations made in testing the hypothesis were made on the basis of both Forms (I and II) combined and averaged. An interesting sidelight on the stability of this prorated "IQ" from the Comprehension and Similarities subtests is
afforded by the fact that, when Vocabulary scores were added into the computations of IQ in the case of the normal group, they made very little difference to the IQ. That is, in this case, three subtest results were combined to give an IQ which, in fact, differed by an average of only 1.7 IQ points from the IQ computed on the basis of only two subtests.

The disturbed couples of the present experiment thus appeared to score very little differently from the normal group in terms of their task efficiency in completing the interaction protocol of the Wechsler Bellevue. At the same time their MMPI scores afforded some clear indications that they should be considered a different population from the normal group if degree or extent of emotional difficulties and pathology is used as the criterion. More light is shed on these group differences by data yet to be presented and further discussion of the point will be postponed until more of the data has been considered.

Hypothesis 1:2 stated that there would be no statistically significant difference between the two groups (normal and disturbed) in terms of the average time taken by couples to complete the interaction protocol. This null hypothesis again could not be rejected (see Table 4).
TABLE 4.—Time taken on the interaction form of the Wechsler Bellevue Comprehension and Similarities subtests by two groups of married couples.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean time (minutes)</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>30</td>
<td>30.433</td>
<td>9.054</td>
</tr>
<tr>
<td>Disturbed</td>
<td>33</td>
<td>30.424</td>
<td>10.314</td>
</tr>
</tbody>
</table>

It seems clear from this table that there is no significant difference between groups in terms of time taken on the task set them \( t = 0.0036 \), and that the variable labeled by Ferreira and Winter (1966) as "Decision Time" did not play a role in the present experiment as it did in theirs. It should be recalled, however, that the task in question was not the same in their study as it was in this. In their case a task specifically designed for their purposes was used rather than a standard psychological test.

However, time was possibly not entirely an irrelevant factor--as the next hypothesis indicates.

**Hypothesis 1:** stated that there would be no statistically significant relationship between "task efficiency" scores on the interaction form of the Wechsler Bellevue and the time taken to complete it. In the normal group, in fact there appeared to be no relationship at all \( r = -.0103 \), but in the disturbed couples
group there was some evidence of a possible relationship 
\( r = -0.3063, p < 0.10 \). The correlation for both groups 
combined was \( -0.1586 \), statistically non-significant.

In interpreting this slight evidence of relationship it should be recalled that a **negative** relationship (such as that found) implies that performance **improved** somewhat the more time was taken, since a **lower** "task efficiency" score implies improved performance. It thus appears that although, as a whole, the disturbed couples group was able to match the normal group in terms of average time taken and average task efficiency, some of the couples in the disturbed group who made better task efficiency scores did so by virtue of taking longer than average time.

**Interaction Testing--Process Scores**

Hypothesis 2 dealt with the qualitative variables of Dominance, Reinforcement, Combination and Emergence, where hypothesis 1 was concerned with the quantitative or product aspects of the data. Each item could be scored for these qualitative variables and Dominance could be scored for either husband or wife. Further, each item could be scored as a **positive**, a **negative**, or a **zero** instance of its kind (Dominance, Reinforcement or whatever), depending if it improved upon, fell short of or made no difference over the better scores of the individual spouses. A hypothetical model of scores for each of the possible cells was set out
on page 50. It would be useful at this point to reproduce this matrix of scores, this time using actual mean scores of the two groups who participated in this study (see Table 5).

**TABLE 5.**—Mean scores of two groups of married couples on 4 process variables derived from Wechsler Bellevue interaction protocols.

<table>
<thead>
<tr>
<th></th>
<th>Plus</th>
<th>Minus</th>
<th>Zero</th>
<th>Total</th>
<th>Total&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CONTROL GROUP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominance (Husb.)</td>
<td>5.80</td>
<td>2.73</td>
<td>5.27</td>
<td>13.73</td>
<td>23.30</td>
</tr>
<tr>
<td>Dominance (Wife)</td>
<td>4.20</td>
<td>1.27</td>
<td>4.00</td>
<td>9.57</td>
<td></td>
</tr>
<tr>
<td>Combination</td>
<td>2.78</td>
<td>1.07</td>
<td>2.87</td>
<td>6.67</td>
<td>10.51</td>
</tr>
<tr>
<td>Emergence</td>
<td>1.80</td>
<td>1.17</td>
<td>0.87</td>
<td>3.84</td>
<td></td>
</tr>
<tr>
<td>Reinforcement</td>
<td>-</td>
<td>-</td>
<td>9.22</td>
<td>9.22</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>14.58</td>
<td>6.24</td>
<td>22.23</td>
<td>43.04</td>
<td></td>
</tr>
<tr>
<td><strong>EXPERIMENTAL GROUP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominance (Husb.)</td>
<td>7.18</td>
<td>2.52</td>
<td>5.67</td>
<td>15.37</td>
<td>26.11</td>
</tr>
<tr>
<td>Dominance (Wife)</td>
<td>4.89</td>
<td>1.76</td>
<td>4.09</td>
<td>10.74</td>
<td></td>
</tr>
<tr>
<td>Combination</td>
<td>1.97</td>
<td>0.89</td>
<td>2.00</td>
<td>4.86</td>
<td>9.44</td>
</tr>
<tr>
<td>Emergence</td>
<td>1.94</td>
<td>1.52</td>
<td>1.12</td>
<td>4.58</td>
<td></td>
</tr>
<tr>
<td>Reinforcement</td>
<td>-</td>
<td>-</td>
<td>7.67</td>
<td>7.67</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>15.98</td>
<td>6.69</td>
<td>20.55</td>
<td>43.22</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>These combined totals were computed in the expectation that certain combinations of cell means might show differences between groups where cell means alone had failed to do so.
On the basis of Bauman and Roman's (1967) findings it was projected that the two groups used in this study would be distinguished in terms of total Reinforcement scores and Negative Emergence scores, and also (on the basis of a small pilot project for the present study) in terms of total Dominance scores (that is, husband and wife scores combined). No differences were predicted for the other variables and cells of Table 5. In fact, as a brief perusal of this table will indicate, in almost every case cell means were very similar from group to group. Even among the three selected for statistical comparison only one entry—that for total Dominance scores—showed a difference approaching significance ($p < .05$). Means, standard deviations and t-ratios for these three variables are given in Table 6.

TABLE 6.—Means, standard deviations and t-ratios for two groups of married couples on 3 process variables derived from Wechsler Bellevue interaction protocols.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>X</th>
<th>S.D.</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Dominance</td>
<td>Normal</td>
<td>23.30</td>
<td>4.53</td>
<td>2.3109a</td>
</tr>
<tr>
<td></td>
<td>Disturbed</td>
<td>26.11</td>
<td>4.87</td>
<td></td>
</tr>
<tr>
<td>Reinforcement</td>
<td>Normal</td>
<td>9.22</td>
<td>3.16</td>
<td>1.8856</td>
</tr>
<tr>
<td></td>
<td>Disturbed</td>
<td>7.67</td>
<td>3.19</td>
<td></td>
</tr>
<tr>
<td>Negative Emergence</td>
<td>Normal</td>
<td>1.17</td>
<td>1.15</td>
<td>1.1429</td>
</tr>
<tr>
<td></td>
<td>Disturbed</td>
<td>1.52</td>
<td>1.23</td>
<td></td>
</tr>
</tbody>
</table>

*$_{p < .05}$.}
In the light of the evidence of Table 6 the following decisions were made with regard to the second group of hypotheses:

**Hypothesis 2:1**, stating that there was no significant difference between groups in terms of total Dominance scores, was rejected.

**Hypothesis 2:2**, stating that there was no significant difference between groups in terms of Reinforcement scores, could not be rejected.

**Hypothesis 2:3**, stating that there was no significant difference between groups in terms of Negative Emergence scores, again could not be rejected.

Of the other variables and among all the remaining cells of Table 7, as well as various possible combinations between cells, none showed any significant difference between groups.

To summarize results of the present study to this point we have found only marginal differences between the two groups of married couples who participated. In terms of the major quantitative criterion used, namely the "task efficiency" score there were no differences. Likewise there was no difference in the amount of time taken to complete the task. There appeared to be a modest relationship among the experimental group between task efficiency score and time taken, a relationship which did not appear in the control group and which may possibly be taken as a
distinguishing element in the performance of the experimental group.

In the second group of hypotheses there were again few indications of differences between groups. The only difference which was statistically significant was in the greater use of Dominance as a mechanism for arriving at a joint answer on the part of the experimental group as compared with the control group.

The Bauman and Roman study, however, offers a second point of comparison with the results of the present study. From the beginning the question was raised whether the present experimental group would achieve scores comparable to those of Bauman and Roman's experimental group and thus replicate their findings. It was always possible that the present experimental group would show major differences in scores from both the present control group and the Bauman and Roman experimental group. It was possible that it might, at the other extreme, be little different from either, even though the two points of comparison were significantly different from each other. Such a finding in fact might support the suggestion of Bauman and Roman that response to the interaction testing procedure might lie along a continuum of "pathology," and approach a first outline for a family typology.

It would be appropriate at this point to consider these findings more closely and to compare the results of
this investigation of hypotheses 1 and 2 with the findings of Bauman and Roman (1968). To do this, one more item needs to be tabulated. Table 7 reproduces the published findings of Bauman and Roman, based on their experimental group alone. For convenience it has been laid out in the same form as Table 5, though it should be noted that Bauman and Roman did not report all the individual cell means of the table. (It should also be noted that their reported totals do not sum to the same grand total in both directions.)

TABLE 7.—Mean scores of 50 married couples on process variables derived from the Wechsler Bellevue interaction test (after Bauman and Roman, 1967).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Plus</th>
<th>Minus</th>
<th>Zero</th>
<th>Total</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominance</td>
<td>16.05</td>
<td></td>
<td></td>
<td>26.18</td>
<td></td>
</tr>
<tr>
<td>(husb.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominance</td>
<td>10.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(wife)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combination</td>
<td>2.74</td>
<td></td>
<td></td>
<td>13.36</td>
<td></td>
</tr>
<tr>
<td>Emergence</td>
<td>10.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reinforcement</td>
<td>6.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10.04</td>
<td>9.78</td>
<td>20.18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A comparison of Tables 5 and 7 in fact affords some interesting comparisons. The Dominance and Reinforcement scores of the disturbed couples' group in this study more
closely resemble those of the Bauman and Roman disturbed couples than they do the normal couples in the present sample. Plus and minus totals, however, liken the disturbed couples' group more to the normal couples' group of the present study than to Bauman and Roman's group. Finally, Emergence and Combination scores show a third trend—in both these cases the present disturbed couples' group have scores which place them between the normal couples and the "pathological" couples of the Bauman and Roman experiment. Thus, if one were to take the normal group of the present study as one extreme and the experimental group of the Bauman and Roman study (where one spouse was receiving inpatient treatment in a psychiatric setting) as the other extreme of a continuum of marital "disturbance," then the present disturbed group could be said to lie somewhere in between, at one time showing scores closer to one extreme at another time showing scores closer to the other, and at yet another time showing scores closer to the middle of the continuum. Table 8 allows for convenient comparison of these suggested trends.
TABLE 8.—Comparison of group mean scores from 3 groups of married couples on 5 process variables derived from Wechsler Bellevue interaction protocols.

<table>
<thead>
<tr>
<th>Group</th>
<th>Dominance</th>
<th>Emergence</th>
<th>Combination</th>
<th>Reinforcement</th>
<th>Minus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (this study)</td>
<td>23.30</td>
<td>3.84</td>
<td>6.67</td>
<td>9.22</td>
<td>6.24</td>
</tr>
<tr>
<td>Disturbed (this study)</td>
<td>26.11**</td>
<td>4.58</td>
<td>4.87</td>
<td>7.67</td>
<td>6.69</td>
</tr>
<tr>
<td>Bauman and Roman</td>
<td>26.18</td>
<td>10.62</td>
<td>2.74</td>
<td>6.28</td>
<td>9.78</td>
</tr>
</tbody>
</table>

The main aim of this part of the present experiment was to delineate the characteristics of the experimental group in contrast to both the control group and the Bauman and Roman experimental group. Another source sheds some light on the same question.

Fox and Weaver (1971, unpublished MS) using virtually the same population as the present experimental group investigated another aspect of the Bauman and Roman study (Bauman and Roman et al., 1967). Bauman had noted that although Interaction IQ did not appear to be a substantial improvement over Husband IQ among their experimental group, yet it could not be assumed that wives made little contribution to the Interaction IQ. In fact, correlations between Husband IQ and Interaction IQ on the one hand and on the other between Wife IQ and Interaction IQ were .814 and .699, respectively. The question whether these individual contributions influenced the Interaction IQ in a
simple additive fashion or in some more complex interactive way was settled by Bauman in favor of a simple additive function when it was found that using only Husband IQs and Wife IQs it was possible to predict Interaction IQ ($r = .88$, close to the theoretically maximum correlation of .92). Further evaluation of the issue showed that husband IQ accounted for 46% of Interaction IQ variance, Wife IQ accounted for 24% and joint contribution accounted for only 8% (which is not significantly different from zero). Fox and Weaver, on the other hand, obtained different results. In the case of their couples, prediction of the Interaction IQ from individual IQs appeared to be a more complex phenomenon than the purely additive one described by Bauman and Roman.

The last section of the analysis of data from the present experiment will be concerned with a rather different question to that of the present part, and it might be as well to attempt a brief summary of findings to date before going on to the final section.

The data gleaned using Bauman and Roman's procedures with interaction testing on the whole proved unproductive, at least in the sense that they indicated few and slight differences in the predicted direction between normal and maritally disturbed couples. From interaction testing alone (without considering the MMPI) one would have to conclude that the disturbed couples' group, with only minor
and unclear exceptions, could well have been drawn from the same population as the normal group.

Conversely, one would have to conclude that the disturbed group of the present study represents a different population from the Bauman and Roman disturbed group, a conclusion which is reinforced by the results of the study by Fox and Weaver (1971) just cited.

Nevertheless, certain other aspects of the total picture offered by the obtained data do not warrant quite so clear-cut a conclusion. The fact is that some of the scores made by the disturbed group in the present study did resemble those of Bauman and Roman's disturbed couples much more than those of the normal population, although differences from the latter's scores were not statistically significant. The more cogent of these suggestive trends were exemplified in Table 8.

The single most provocative finding of this study, however, was the indication afforded by MMPI testing, quite clearly reinforcing what was only implicit and suggestive in Table 8, that the disturbed group of the present study did indeed represent a different population to the normal group. The details of the difference have already been set out. What makes this finding provocative is that it shows the MMPI as apparently more sensitive to group differences than the interaction testing procedure, at least to
measurable group differences. It is unfortunate that MMPI profiles for the Bauman and Roman disturbed couples are not available. Clinical experience suggests that the hospitalized spouse would almost certainly show an elevated MMPI profile at admission, and thus liken the group to the disturbed couples of the present study. But what of the non-patient spouse's MMPI in the case of the Bauman and Roman couples? It is virtually impossible to assess whether the present experiment's disturbed couples resemble their counterparts in Bauman and Roman's study in terms of MMPI scores. Nor was it possible to find any consistent pattern within the scores of the present disturbed group. Of the 5 couples, for example, with no elevations on the MMPI clinical scales, 4 had "task efficiency" scores among the poorest of the whole group. And on the other hand, some couples showing quite acute emotional stress on the MMPI had among the best "task efficiency" scores. Whatever degree or kind of disturbance in their marriages experienced by these couples, neither the MMPI nor interaction testing was refined enough to register it unambiguously, though both seemed to reflect it in some general way. Interaction testing with the Wechsler Bellevue, in particular, showed low power of discrimination in distinguishing between the normal and the maritally disturbed groups of the present study.
Factors Affecting Dominance Scores

The final section of this paper deals with a more detailed analysis of the single variable of Dominance. Dominance is scored on the joint or Interaction protocol when the response of one partner from his or her individual protocol is used to the exclusion of the other's. Bauman and Roman (1966) found that multiple factors seemed to determine this phenomenon. In the first place husbands' responses tended to dominate over wives', the more intelligent partner's over the less intelligent, the non-patient's over the patient's, and the recorder's (the one who actually wrote down the joint answers) over the non-recorder's.

In the present case there was no distinction between a diagnosed "patient" and a non-patient in each couple. They had presented themselves as needing conjoint therapy. Further, a preliminary series of separate t-tests pairing the factors isolated by Bauman and Roman indicated that "recorders mean scores on the Dominance dimension did not differ significantly from non-recorders." This last factor was therefore eliminated from further consideration.

On the other hand it has already been noted that Dominance scores were significantly higher among the experimental couples than among the controls. And again, reasoning from Wechsler's (1958) observation that the Comprehension subtest seemed consistently to favor men
while the Similarities subtest favored women, it was hypothesized that the nature of the subtest would also have an effect on Dominance scores. In the present experiment, therefore, an analysis of variance was performed on Dominance scores which omitted the Bauman and Roman factors of "patient" vs. "non-patient" and "recorder" vs."non-recorder," but which included subtest differences and also group differences, which later seemed (contrary to Bauman and Roman's findings) to exert some influence in the present experiment. The independent variables in this analysis of variance were group (factor a), relative intelligence (factor b), and sex (factor c). The dependent variable was Dominance score, but analyzed separately for each of the two subtests as well as for both together. Results are set out in Tables 9, 10 and 11.

This analysis of variance reveals some consistent trends. In particular it points up the value of examining the results for each subtest separately.

Hypothesis 2:1 had already revealed that the experimental group tended to use Dominance as a mechanism for arriving at joint answers more than did the control group. Analysis of variance showed that this difference was apparently due to group differences on subtest 1 (Comprehension) alone. An inspection of cell means (Table 10) shows negligible differences on subtest 2 (Similarities).
TABLE 9.—Analysis of variance of mean Dominance scores from two groups of married couples on the Wechsler Bellevue interaction protocol.

a) Cell means (Dominance) for group vs intelligence status vs sex on subtest Comprehension of the W-B.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>cl&lt;sup&gt;a&lt;/sup&gt;</th>
<th>c2&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1a b1</td>
<td>14</td>
<td>6.57</td>
<td>3.57</td>
</tr>
<tr>
<td>a1a b2</td>
<td>16</td>
<td>6.00</td>
<td>4.12</td>
</tr>
<tr>
<td>a2a b1</td>
<td>15</td>
<td>9.20</td>
<td>3.87</td>
</tr>
<tr>
<td>a2a b2</td>
<td>17</td>
<td>6.47</td>
<td>4.88</td>
</tr>
</tbody>
</table>

b) Analysis of variance for group vs intelligence status vs sex on subtest Comprehension of the W-B.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1</td>
<td>33.3697</td>
<td>33.3697</td>
<td>8.0933</td>
<td>&lt;.02</td>
</tr>
<tr>
<td>IQ status</td>
<td>1</td>
<td>5.8372</td>
<td>5.8372</td>
<td>1.4157</td>
<td>&lt;.02</td>
</tr>
<tr>
<td>Group x IQ status</td>
<td>1</td>
<td>5.5734</td>
<td>5.5734</td>
<td>1.3517</td>
<td>&lt;.02</td>
</tr>
<tr>
<td>Within</td>
<td>58</td>
<td>239.1401</td>
<td>4.1231</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>1</td>
<td>268.4912</td>
<td>268.4912</td>
<td>33.1613</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sex x group</td>
<td>1</td>
<td>8.0246</td>
<td>8.0246</td>
<td>0.9911</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Sex x IQ status</td>
<td>1</td>
<td>45.5455</td>
<td>45.5455</td>
<td>5.6253</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Sex x IQ status x group</td>
<td>1</td>
<td>13.2355</td>
<td>13.2355</td>
<td>1.6347</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Within</td>
<td>58</td>
<td>469.6006</td>
<td>8.0965</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>a1 = control group  
a2 = disturbed group  
b1 = husband more intelligent  
b2 = husband less intelligent  
c1 = husband  
c2 = wife
TABLE 10.—Analysis of variance of mean Dominance scores from two groups of married couples on the Wechsler Bellevue interaction protocol.

a) Cell means (Dominance) for group vs intelligence status vs sex on subtest Similarities of the W-B.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>cl&lt;sup&gt;a&lt;/sup&gt;</th>
<th>c2&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14</td>
<td>8.36</td>
<td>5.00</td>
</tr>
<tr>
<td>b1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>16</td>
<td>6.80</td>
<td>6.19</td>
</tr>
<tr>
<td>b2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>15</td>
<td>10.40</td>
<td>5.33</td>
</tr>
<tr>
<td>a2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>17</td>
<td>5.71</td>
<td>6.76</td>
</tr>
</tbody>
</table>

b) Analysis of variance for group vs intelligence status vs sex on subtest Similarities of the W-B.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>(61)</td>
<td>6.5283</td>
<td>6.5283</td>
<td>0.8989</td>
<td></td>
</tr>
<tr>
<td>IQ status</td>
<td>1</td>
<td>25.2679</td>
<td>25.2679</td>
<td>3.4795</td>
<td></td>
</tr>
<tr>
<td>Group x IQ status</td>
<td>1</td>
<td>16.2174</td>
<td>16.2174</td>
<td>2.2332</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>58</td>
<td>421.1916</td>
<td>7.2619</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>(62)</td>
<td>8.0000</td>
<td>8.0000</td>
<td>0.5866</td>
<td></td>
</tr>
<tr>
<td>Sex x group</td>
<td>1</td>
<td>0.0030</td>
<td>0.0030</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Sex x IQ status</td>
<td>1</td>
<td>151.3696</td>
<td>151.3686</td>
<td>11.1002</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Sex x IQ status x group</td>
<td>1</td>
<td>36.0415</td>
<td>36.0415</td>
<td>2.6430</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>58</td>
<td>790.9195</td>
<td>13.6365</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup><sup>a1</sup> = normal group  
<sup>a2</sup> = disturbed group  
<sup>b1</sup> = husband more intelligent  
<sup>b2</sup> = wife more intelligent  
<sup>c1</sup> = husband  
<sup>c2</sup> = wife
TABLE 11.---Analysis of variance of mean Dominance scores from two groups of married couples on the Wechsler Bellevue Interaction protocol.

a) Cell means (Dominance) for group vs intelligence status vs sex on both subtests Comprehension and Similarities of the Wechsler Bellevue.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>cl&lt;sup&gt;a&lt;/sup&gt;</th>
<th>c&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14</td>
<td>14.93</td>
<td>8.57</td>
</tr>
<tr>
<td>b2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>16</td>
<td>12.81</td>
<td>10.31</td>
</tr>
<tr>
<td>a2&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>15</td>
<td>19.60</td>
<td>9.20</td>
</tr>
<tr>
<td>b2</td>
<td>17</td>
<td>12.18</td>
<td>11.65</td>
</tr>
</tbody>
</table>

b) Analysis of variance for group vs intelligence status vs sex on both subtests Comprehension and Similarities of the Wechsler Bellevue.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1</td>
<td>69.6457</td>
<td>69.6457</td>
<td>6.9879</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>IQ status</td>
<td>1</td>
<td>55.1884</td>
<td>55.1884</td>
<td>5.5373</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Group x IQ status</td>
<td>1</td>
<td>40.6277</td>
<td>40.6277</td>
<td>4.0764</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>58</td>
<td>578.0603</td>
<td>9.9665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>1</td>
<td>755.1893</td>
<td>755.1893</td>
<td>27.4352</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sex x group</td>
<td>1</td>
<td>8.2653</td>
<td>8.2653</td>
<td>0.3002</td>
<td></td>
</tr>
<tr>
<td>Sex x IQ status</td>
<td>1</td>
<td>363.5052</td>
<td>363.5052</td>
<td>13.2057</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sex x IQ status x group</td>
<td>1</td>
<td>69.6472</td>
<td>69.6472</td>
<td>2.5302</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>58</td>
<td>1596.5249</td>
<td>27.5262</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup><sub>1</sub> = normal group
<sup>a</sup><sub>2</sub> = disturbed group
<sup>b</sup><sub>1</sub> = husband more intelligent
<sup>b</sup><sub>2</sub> = wife more intelligent
<sup>c</sup><sub>1</sub> = husband
<sup>c</sup><sub>2</sub> = wife
Similarly, the statistically significant preponderance of husband Dominance over wife Dominance means was paralleled by the highly significant husband Dominance on subtest 1 alone. Subtest 2 did indeed underline the same trend for husbands' responses to outnumber wives' on the Dominance dimension. But the difference in this case was not significant. This latter finding appears to offer strong corroboration of the research hypothesis which prompted this aspect of the experiment. It had been noted that the various subtests of Wechsler's intelligence scale consistently favor one sex or the other. It had also been noted that Dauman and Roman's (1966) analysis of the data from their experiment indicated a strong tendency for husband's individual responses to dominate on the joint protocol. On this basis it was suggested that perhaps husband's responses would dominate much less on any particular subtest which was considered to favor women. This indeed seems to have been the case. On subtest 1 (Comprehension) which usually favors men husband Dominance scores far outweighed wife Dominance scores ($p < .001$). On subtest 2, however, where women usually do better, the difference was still in favor of the husbands but it was not significant. In fact one subgroup of wives actually outscored their husbands on Dominance for subtest 2.

This particular subgroup was the group where the wife had a higher IQ, suggesting that IQ status too played a role
in determining Dominance scores. This, again, was totally in accord with the findings of Bauman and Roman. In the present study, of 24 couples (taken from both groups combined) where the husband was the more intelligent, the husband's Dominance score was greater than his wife's in all but one case. Where the wife was the more intelligent on the other hand (29 cases) her Dominance score was greater than her husband's in only 11 instances. This would appear to be a clearly interactive phenomenon. There was an overall tendency for husbands to dominate, but where wives were the more intelligent this tendency was modified. As has already been pointed out, the same tendency was also modified by the particular subtest used. Thus the interaction of sex vs intelligence was more marked \((p < .01)\) on subtest 2 than it was on subtest 1, where the tendency for the more intelligent partner to dominate was apparently weakened by this subtest's partiality towards men as opposed to women \((p < .05)\).

It should be noted, finally, that intelligence status alone (and not in interaction) appeared to be a relevant factor influencing Dominance scores only on the combined means of subtests 1 and 2 (see Table 11). It would seem that this was not due to one subtest rather than the other, but was rather the cumulative manifestation of a trend which was evident in each of the subtests taken separately but was not significant.
One of the main research hypotheses of this experiment was that the nature of the subtest used would influence Dominance scores on the interaction form of the Wechsler Bellevue. As the preceding discussion has sought to bring out, analysis of variance of couples' Dominance scores on the Comprehension and Similarities subtests lends support to this hypothesis. Further confirmation was sought through an analysis of variance of scores on the Vocabulary subtest which again favors women over men, the differential in this case being usually one of the largest among all the subtests. The expectation, therefore, would be that the trends noted in the analysis of scores from subtest 2 would be at least equally and perhaps more manifest in Vocabulary scores. Results are given in Table 12.

The fact that none of the F-ratios in Table 12 is statistically significant is exactly what one would have projected on the basis of previous results, and confirms that the nature of the subtest influences Dominance scores. A comparison of the cell means from Table 10 (a: Similarities) and Table 12 (a: Vocabulary) reveals an almost identical pattern. In both cases it appears that the most general tendency of all (the tendency of husband responses to dominate) is mitigated not only by the cases where the wife happens to be the more intelligent but also by the nature of the subtest which favors women.
TABLE 12.—Analysis of variance of mean Dominance scores of 30 married couples (control group) on the Wechsler Bellevue interaction protocol, using Vocabulary subtest only.

a) Cell means (Dominance) for intelligence status vs sex on the Vocabulary subtest of the Wechsler Bellevue.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>b1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>b2&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14</td>
<td>11.71</td>
<td>6.29</td>
</tr>
<tr>
<td>a2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>16</td>
<td>8.44</td>
<td>8.19</td>
</tr>
</tbody>
</table>

b) Analysis of variance for intelligence status vs sex on the Vocabulary subtest of the Wechsler Bellevue.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ status</td>
<td>1</td>
<td>7.0082</td>
<td>7.0082</td>
<td>1.6234</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>28</td>
<td>120.8750</td>
<td>4.3169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>1</td>
<td>120.0476</td>
<td>120.0476</td>
<td>4.1334</td>
<td></td>
</tr>
<tr>
<td>Sex x IQ status</td>
<td>1</td>
<td>99.8101</td>
<td>99.8101</td>
<td>3.4365</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>28</td>
<td>813.2144</td>
<td>29.0433</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>a1 = husband more intelligent  
<sup>a</sup>a2 = wife more intelligent  
<sup>a</sup>b1 = husband  
<sup>a</sup>b2 = wife
The present study has thus demonstrated that the variables isolated by Bauman and Roman as having an influence on Dominance scores—namely sex and intelligence—also had an influence on Dominance in the present case. It has further demonstrated that group differences (contrary to Bauman and Roman) have some effect on Dominance scores. This effect, however, was traced to one particular subtest. The influence of individual subtests on Dominance scores was further corroborated by the separate analyses of variance which were performed on three subtests separately.

On the basis of these results, Hypothesis 3, that Dominance scores will not be determined by sex, intelligence or marital disturbance, was rejected in all its parts.

Discussion

There were two principal thrusts behind the experiment which has been described. The first was an exploration of the proposition put forward by Roman and Bauman (1967) that couples exhibiting varying degrees of personal and marital pathology would possibly be found to lie on a continuum—a continuum which would be mirrored in their different modes of response to the interactional form of parts of the Wechsler Bellevue intelligence scale. On the whole, the present experiment lent only modest support to
this proposal. On the grosser measures of difference between the experimental and control groups only two items seemed to have any effect, namely the greater number of Dominance scores among the experimental group and the relationship, in the same group, between time taken and task efficiency, a relationship which did not seem to exist in the control group. Neither of these two items, however, reproduced Bauman and Roman's findings. They did not report any measure of time taken, and differences between groups on Dominance scores were non-significant. It should be noted also that the difference in Dominance scores discovered in the present study seemed to be due not to the fact that the experimental group exceeded the Dominance scores of the Bauman and Roman group (in fact the means were almost identical) but rather because the normal or control group had apparently significantly lower scores than Bauman and Roman's normal group. One possibility which comes to mind to explain this phenomenon is that scoring criteria were differently interpreted in the two experiments.

Possibly the most disappointing aspect of the results of this study, in view of the hope that some clear guidelines towards the establishing of a couples' typology might emerge, was the fact that even though the interaction testing procedure showed only negligible differences between the experimental and the control group, the MMPI seemed to
have the sensitivity to distinguish the two groups fairly sharply.

However, it must be recognized that this experiment and the one by Bauman and Roman represent soundings at two convenient but widely differing points and it should not be too surprising that a straight line cannot immediately be drawn between the two points. It should also be said again that present results were not wholly without indications of the sort of continuum Bauman and Roman had in mind. As Table 8 made clear, there were areas where the experimental group used in the present study scored in quite similar fashion to its own control group, and areas again where it came much closer to the reported scores of Bauman and Roman's experimental group. It may be that the concept of a continuum will have to be broken down more carefully. It may be, for example, that total negative scores, total Emergence scores or total Combination scores will prove to be more sensitive to differences in pathology than more general measures (such as task efficiency). There is some slight evidence that these three aspects of the scoring table could more quickly show up group differences than would any overall summary score. Indeed it is possible that a statistical comparison of scores on these three factors between the present control group and the Bauman and Roman experimental group might show some significant differences.
There are two outstanding problems, however, if these more fragmentary aspects of the data are to be explored more thoroughly. The first is that with only 44 items (Comprehension and Similarities subtests combined) and fifteen cells to which each response can be assigned, cell frequencies are in some cases quite small, not to mention the fact that in both this and in Bauman and Roman's experiment over half the responses (irrespective of group) had to be assigned just to the Dominance cells. Not many of the Wechsler subtests can be appropriately used in interaction testing, but the Vocabulary subtest was successfully used in this experiment and its addition almost doubled the item pool. Doubtless a diligent item analysis of the results of both this and Bauman and Roman's experiment would allow a certain pruning of items which contributed little and it might be that the three subtests together would provide an eventual item pool of something like 65 units.

There is, of course, nothing sacred about the Wechsler items as such. It may be that a more effective item pool could be devised using a different type of item altogether—effective in the sense that it would show greater sensitivity with differently composed groups of people than apparently did the present item pool.

The other problem which would need to be solved if analysis of data were to be based on subtotals rather than
overall totals would be that of scoring criteria. All one can do at the moment is to follow the ones proposed by Bauman and Roman, in the interests of standardized research. There are at present, however, some problems with these criteria and Bauman and Roman themselves may not feel they have yet hit on an ideal system. Certainly the criteria are more difficult to apply than appears at first blush.

Logically it would seem that there are only four possibilities where two people are trying to arrive at a joint solution to a problem for which they have already found their individual solutions. The first would be that both hit on the same answer and readily agree to use that answer in their consensus response. This is Bauman and Roman's "Reinforcement." In the event that the two individuals did not have precisely the same answer it would seem that they could either (1) agree to drop one of their individual responses ("Dominance"); (2) combine elements of both answers ("Combination"); or (3) solve their differences by coming up with something entirely new ("Emergence"). Yet there are cases which do not fit comfortably under any rubric. Roman and Bauman (1964) in their revised administration and scoring manual give two excellent examples, as follows: to the question "Why are houses built of brick rather than wood? these hypothetical responses are given:
<table>
<thead>
<tr>
<th>Husband A</th>
<th>Wife A</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;It's cooler&quot;</td>
<td>&quot;lasts longer&quot;</td>
<td>&quot;lasts longer&quot;</td>
</tr>
<tr>
<td>&quot;fire protection&quot;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Husband B</th>
<th>Wife B</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Lasts longer&quot;</td>
<td>&quot;lasts longer&quot;</td>
<td>&quot;lasts longer&quot;</td>
</tr>
<tr>
<td>&quot;fire protection&quot;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first example is scored by Roman and Bauman as Dominance (wife) "insufficient," and the second as Dominance (husband) "insufficient." Both instances would also be scored as "negative" interactions in that in each case the wife's response was scored 2 and the interaction response was scored 1. It is hard to see the first case as an example of "negative wife Dominance." The second case is indeed an example of partial Dominance on the husband's part—partially because it did not entail the wife sacrificing her reply completely. Bauman and Roman are right in placing such cases under a distinguishing notation such as "insufficient." But then are such instances to be included in totals, thus becoming indistinguishable again from responses which more fully meet the criteria for Dominance? It might be better to devise another category for this type of response, which in fact appeared sufficiently often in the present experiment to have made it worthwhile looking for a separate category. Both cases are in fact examples of crucial omissions—and Omission might be a better title. Again it might be considered that the interaction responses represented not so much a combination of individual
responses but more of a "compromise." Possibly the title Compromise might be more satisfactory.

A parallel dissatisfaction with the present scoring system lies in the problem of merely verbal Dominance. If one partner says, in answer to the question "Why must we pay taxes?, "to run the government" and the other says "to support the government," does it matter whose actual words are used? Should this example be scored for and included among Dominance totals? Is it not more an example of Reinforcement (at the level of ideas)? Should it be scored for both Dominance and Reinforcement? Can it rightly be scored for Dominance at all? There are couples where one partner seems to insist on his or her actual words even when the ideas expressed are the same, and this may well be clinically important. But to consistently score such a phenomenon for Dominance would surely over-inflate the Dominance totals. Bauman and Roman do not say whether such instances should be included in Dominance totals. It is another point which requires clarification.

The last part of this experiment was concerned with Dominance alone and again the importance of clear criteria in scoring will be of considerable help in comparing one piece of research with another.

In this part of the experiment, results entirely supported the findings of Bauman and Roman (1966). The three variables selected for analysis of variance (group,
intelligence and sex) all showed main effects, and intelligence x sex also showed a first-order interaction effect. It seemed that there was an overall tendency for husband's responses to be used on the interaction protocol in preference to wife's. The tendency seemed to be suppressed somewhat in the normal group and it seemed to be mitigated further when the subtest favored women or when the wife was the more intelligent of the two partners, or both. But the tendency was only suppressed completely (and then just barely) in the case of the small group of wives in the experimental group who happened to be the more intelligent partner and only in the case of subtest 2. In this one case, wives' mean Dominance score narrowly outweighed the husbands' mean Dominance score.

As Bauman and Roman (1966) note, it is hard to say how much of this dominant behavior on the part of husbands is related to role and status expectations of a purely cultural kind and how far to psychobiological sex differences in aggressive or initiatory behavior. They include preponderant husband Dominance among "non-rational" influences on Dominance scores and the impression was given by the present experiment that spouses tended to "trade off" responses and alternate Dominance scores in a way that was sometimes unproductive.

In conclusion it should be noted that merely listening to the couples who participated in this experiment,
both normal and experimental, seemed to offer a rich view of clinical data which the scoring procedures do not, of course, attempt to measure. Interaction testing is a fairly new phenomenon and should not be overrated. But a combination of standardized scoring procedures such as those evolved by Bauman and Roman with the clinical or impressionistic data which a tape-recording of the interaction testing session would furnish might prove to be a powerful tool of diagnosis and a potent means of confronting a couple in the therapy session itself with an example of unproductive interaction which they could acknowledge and whose occurrence a sufficient number of times would enable the therapist to establish a comparison between the couple and different populations of couples.

Interaction testing, as has been said is still in its early years, and there are many fruitful directions in which it can be explored. Clearly, cross-cultural differences and differently composed dyads both within the family and without could be examined in terms of the score patterns they produce. In terms of the results of the present experiment a fruitful approach at this point might well be to establish a priori (on the basis of an accepted personality test) some continuum of degree of pathology, deriving the total population from any source--hospitalized or out-patient clients, normal population, couples with one, with two or with no acutely disturbed member. The next step
would be to construct a number of cells such as "psychotic," "neurotic" and "normal," subdividing them further according to whether one or both spouses should be included. With some such classificatory system for ranking groups of couples in order of degree of pathology it might be possible to parallel such a ranking with rankings of scores on the Wechsler Bellevue interaction testing procedure. It might be that more specific results could be obtained in this way than by the method adopted in the present experiment which was simply to divide couples into groups on the basis of their involvement in out-patient conjoint therapy.

In view of the present study it should further be kept in mind that care in constructing adequate groups will also have to take into account the fact that Dominance scores are affected by several factors which need to be controlled. A group of couples which included, for example, an unusual proportion of cases where the wife was the more intelligent would not be expected to show the same results as a group where most of the husbands happened to be the more intelligent. This and the Bauman and Roman experiment have indicated that factors which might affect Dominance scores include patient-status of one partner, intelligence status of one partner, psychological variables (such as those measured by MMPI), and specific item pool used. Almost certainly other factors with a potential influence on Dominance scores (perhaps socio-economic
level) will be isolated. Experimental control and manipulation of these variables will help to shed more light on the possibilities and limitations of interaction testing.
CHAPTER V

SUMMARY

The starting point of the present study lies in a series of studies using a comparatively new diagnostic technique, namely "interaction testing." The basic format of this approach involves the administration of an instrument (usually but not necessarily one of the standard psychometric tests) to two or more individuals as a single entity. That is, all the members of the particular group being tested are asked to agree on a single set of joint, consensus answers to the test.

A variety of tests have been used within the same format, perhaps the most popular being the Rorschach; a variety of methods have been applied in interpreting results, perhaps the most common being an analysis of interaction on the Bales (1950) model of the group's discussion; finally, the format has been used with a variety of groups, both in terms of size and of composition.

The present study uses the Comprehension and Similarities subtest of the Wechsler Bellevue intelligence scale; analysis of the data focusses exclusively on the end product of the interaction, namely the joint protocol.
itself; and Ss are husband-wife dyads.

The researchers responsible for initiating this particular line of inquiry, Gerald Bauman and Melvin Roman, have published a series of papers during the 1960s detailing their experiment with a group of married couples (N = 50), one of whom in each case was currently hospitalized for psychiatric reasons. To better understand their results and those of the present experiment it will be necessary to understand something of the specific methodology they employed, a methodology followed in the present study.

First, the two Wechsler subtests, Comprehension and Similarities (Form I and Form II) are administered to each spouse separately. Their protocols are scored in the traditional way, each item being worth 2, 1, or 0 points according to Wechsler's criteria. The raw scores are then converted to scale scores, and the scale scores are prorated to give a Verbal IQ, all again according to Wechsler's standard procedures. Thus each partner obtains an individual IQ.

Upon completion of their individual protocols, the couple are invited to complete an identical protocol as if they were a single individual, that is, by agreeing to a joint answer for each item. The protocol is scored as though from a single individual and again according to standard procedures. An "Interaction IQ" is thus obtained.
A fourth protocol is then derived (artificially) by taking each item in turn and scoring only for the better answer each time from one or other of the two original individual protocols. This protocol is also treated as if it came from an individual and scored in the standard manner. It represents the best the couple might have done in interaction had they selected the better individual answer for each item. It is called the "Potential IQ."

In almost all cases there will be a gap between the Potential IQ (the best a couple might have done) and the Interaction IQ (what they actually did, in interaction). Bauman and Roman called this gap (Potential IQ minus Interaction IQ) the "task efficiency" score. They found that, averaged over groups, this task efficiency score distinguished at a statistically significant level between the disturbed group and the normal group used as a control ($t = 4.18, p < .001$). They also instituted a more detailed item analysis of scores (a procedure followed in the present experiment) as follows.

First, each item of the Interaction protocol can be assigned a "positive," "negative" or "zero" quality, depending whether it equalled or surpassed the better individual score (that is, 2, 1 or 0 according to Wechsler's criteria), in which case it was a positive response, whether it fell short of the better individual response, in which case it is scored negative, or whether
the joint response happened to have the same value as both individual responses, in which case it was a zero response. Bauman and Roman's finding that the normal group had a significantly better task efficiency score was paralleled by the finding that the disturbed group had significantly more "negative" responses than the normal group. That is, there were significantly more instances on their interaction protocol where they had used an answer inferior to that given by one or both on their original individual protocols ($t = 5.33, p < .0005$).

These two findings, of course, are related in that they are two different assessments of the same thing, namely a couple's ability to maximize their combined resources in completing this particular task. It was Bauman and Roman's hope that failure or success in meeting this task would prove a sensitive enough reflector of marital discomfort that it would be able to register a continuum of degree of distress and mark a first step towards establishing a couples' typology along this dimension.

The analysis of results, as described so far, could be termed "quantitative" analysis or product analysis. It is also possible to make a "qualitative" or process analysis of the data. Again, this involves item analysis.

This time, each item is scored for (1) Dominance, where one partner's answer is used in interaction and the
other's is sacrificed; for (2) **Combination**, where elements of both partner's original responses are used in interaction; for (3) **Emergence**, where an entirely new element is used in interaction; for (4) **Reinforcement**, where the individual answers happened to coincide and this response was adopted also as the joint response.

The Dominance scores can be further distinguished into Husband Dominance and Wife Dominance scores. Each of the 5 resulting process variables (Husband Dominance, Wife Dominance, Combination, Emergence and Reinforcement) can again be designated as either a "positive," a "negative" or a "zero" interaction, thus yielding a total of 5 x 3 cells into which each interaction response can be placed. In comparing cell frequencies from group to group, as well as various totals and sub-totals, Bauman and Roman found that only two showed statistically significant differences. They found that Reinforcement occurred more often in the normal group and "negative" Emergence more often in the disturbed group (p < .001 in both cases).

Bauman and Roman thus opened up the promising prospect of a simple, short and non-threatening testing procedure which could distinguish between normal couples and disturbed couples. The question remained, however, whether the disturbed couples of their study were representative of disturbed couples in general. It was toward answering this question that the present experiment was shaped. In
particular it seemed that application of Bauman and Roman's procedures to a group of couples who had applied for conjoint therapy at an outpatient marriage clinic would afford valuable information about a population which could be called maritally "disturbed" but which likely was not disturbed in the same sense that Bauman and Roman's couples were, where one or the other was receiving psychiatric in-patient treatment.

The question whether a differently composed group of maritally disturbed couples would respond to the Wechsler Bellevue interaction testing procedures in the same way as had Bauman and Roman's disturbed group seemed justified by the results of the present study. In this case a sample of couples was drawn from the population at large (N = 30) and another sample from couples who had applied for joint counseling at a university hospital family outpatient clinic (N = 33). The MMPI distinguished between the two groups rather more clearly than had been anticipated. Where only two members of the normal group gave signs of acute emotional disturbance (in terms of considerable or multiple elevations on the MMPI clinical scales) no less than 30 from the disturbed group gave a profile of this type. Conversely, where only 5 couples of the disturbed group gave profiles completely within the normal range of the clinical scales, there were 19 from the normal group who did so.
Interaction testing with the Wechsler Bellevue, on the other hand, failed to distinguish quite so sharply between the two groups. Testing procedures and scoring were identical with those of Bauman and Roman (just described), and in line with their findings it was hypothesized that the two groups would be distinguished in terms of their "task efficiency" score, and in terms of the Reinforcement and "negative" Emergence scores. In line with the findings of Ferreira and Winter (1965) it was also hypothesized that the disturbed group would take longer to complete the interaction protocol than would the normal group, and (on the basis of a pilot study for the present project) that Dominance scores would distinguish at a statistically significant level between the two groups. Results were as follows:

Task efficiency score failed to distinguish the present disturbed group from the normal group (t = 0.215, p < .40). Time taken to complete the interaction protocol (a variable which another team of researchers (Ferreira and Winter, 1965) had found to characterize malfunctioning families) did not distinguish between the two groups in this present study (t = .0036). A slight relationship in the case of the disturbed couples between time taken and task efficiency (r = -.3063, p < .10) was not found among the normal couples. This was the only difference found
between the two groups in terms of various aspects of the "task efficiency" score.

Results were similarly patterned in the case of the 5 process variables (Husband Dominance, Wife Dominance, Combination, Emergence and Reinforcement) isolated by Bauman and Roman. Their results were again not replicated in the present experiment. In the present case there were no statistically significant differences between groups as there were in their case in terms of Reinforcement scores, total negative outcomes, or negative Emergence scores. On the other hand the present study did show a difference between groups in the one case of total Dominance scores ($t = 2.3109, p < .05$), a difference, however which implied varying group means between the two normal groups (the present one and Bauman and Roman's), and suggested the possibility that differences in scoring techniques might be responsible for the difference in totals.

At this point results of the present study would have to be termed disappointing in that they had failed to pinpoint any clearcut interaction differences between the normal and disturbed groups. However, trends such as those suggested by Bauman and Roman were visible, though not statistically significant. The disturbed group did indeed in some cases match Bauman and Roman's couples more nearly than normal couples. In the case of other variables (such
as total Combination, Emergence and Reinforcement scores) the scores of the disturbed group seemed to lie somewhere in between those of the normal group at one extreme and those of Bauman and Roman's disturbed group at the other. There was thus some evidence that a continuum of scores might eventually be disclosed which would enable a researcher to place different couples into different categories according to degree of disturbance. It was clear, however, that such a continuum would be established in all probability not so much through grosser or more general measures such as "task efficiency" but through careful analysis of the separate means of the different process variables. It seemed likely too, on this assumption, that the testing procedure would require a somewhat larger item pool if it were to acquire the power to generate statistical differences among cell means. With a present item pool of 44 items and 15 cells to which each could theoretically be assigned it seemed that cell means would likely be too small for true statistical comparison. In the present experiment (though only in the case of the normal group) the Wechsler Vocabulary subtest was added to the item pool and proved to be suitable material for the interaction testing procedure as well as almost doubling the item pool.

Perhaps the most striking aspect of the results of the present study came from MMPI scores. The results here
were somewhat unexpected. It had been anticipated that the normal couples would almost all show profiles within the normal range or at least that deviations from the normal profile would be few and slight. This expectation was borne out by the results. It had also been expected that the disturbed group would likewise reflect a cross-section of the normal population—a part of the normal population which happened, currently, to be experiencing stress in marriage. Such expectations were, however, shown to be unwarranted. Clinically, this proved to be a group where the majority of individuals showed deviations from the normal profile, and in virtually half of the cases such deviations indicated acute emotional disturbance. It thus appeared that the slight trends indicated by interaction testing, pointing to some differences between the disturbed group and the normal group of the present study, were validated in a general way by the MMPI scores, which pointed strongly to the conclusion that the two groups should be considered as representing different populations in terms of personal or emotional disturbance. The lack of power or sensitivity of the interaction procedure to reflect this difference was considered to be disappointing. Again, it seemed that the composition of the item pool might have to be reconsidered if the interaction testing approach were to gain in such power and sensitivity.
One last aspect of both the Bauman and Roman study and the present one lies in a more detailed analysis of the single process variable, Dominance. In examining a variety of factors which seemed to have a possible influence on Dominance scores, Bauman and Roman (1966) showed clearly that these scores were strongly affected by sex (husbands' responses dominated much more than wives'), by intelligence status (the higher IQ partner's responses dominated more than the lower IQ partner), by recorder status (that is, the responses of the spouse who happened to be the recorder of their joint responses tended to prevail over that of the non-recorder) and by current hospital status (the non-patient's answers dominated over the patient's).

A similar analysis of variance was performed in the present case though with some exceptions. In the first place there was no "patient" distinct from a "non-patient" in the present sample. Secondly, a preliminary examination of the different variables through a series of t-tests seemed to indicate that recorder status played only a small part in the present Dominance scores. It too (recorder status) was eliminated from the analysis. At the same time, preliminary analyses had also indicated that group differences might play a role here where they had not had any apparent effect in the Bauman and Roman study. It will be recalled that Dominance scores were significantly higher in the disturbed group than they were in the normal.
Lastly, an original element of the present study lay in its analysis of Dominance scores separately by subtest. There was reason to think that different subtests might influence Dominance scores in differing ways. Wechsler's (1958) work seemed to indicate that the Comprehension subtest consistently favored men over women, while the Similarities subtest tended to favor the women. It was reasonable to suppose that the slight edge which women might be expected to have over men in a subtest such as Similarities might diminish the tendency of husband responses to dominate. As a further test of this hypothesis the Vocabulary subtest was administered to the normal group, in addition to the other two.

Thus, the independent variables in the present analysis of variance were group, intelligence status and sex; and the dependent variable was Dominance score, analyzed separately for each subtest as well as for both subtests together. In the case of the normal group, a third subtest was added to the analysis.

Results entirely supported both Bauman and Roman's findings (in so far as their study was replicated here) and also the supposition that the nature of the subtest would influence Dominance scores. There was a clearly visible overall tendency for husbands' responses to dominate over wives' as there was in the Bauman and Roman experiment (in the present case $F = 27.4352, p < .001$). However, it
appeared that this prevailing Dominance on the part of husbands was due almost entirely to Dominance on subtest 1 alone (F = 33.1613, p < .001). Husband Dominance on subtest 2 (Similarities) where the test seemed likely to favor the wives showed only a slight edge in terms of husband Dominance. Again, on the Vocabulary subtest (again apparently favoring women) the husband Dominance scores, although again tending to outnumber wives' Dominance scores did not do so to a significant extent. It seems not unreasonable, therefore, to conclude that though husbands continued to outscore wives in Dominance the effect of the nature of the particular subtest was such as to mitigate this general tendency where the subtest appears to favor women's scores.

Secondly, the difference between groups already reported in the case of total Dominance scores was reflected in the analysis of variance (F = 6.9879, p < .05). Again, however, separate analyses of subtest results indicates that this was largely due to the results of subtest 1 (Comprehension) alone (F = 8.0933, p < .02). It seems that, once more, an interactive effect is influencing Dominance scores. In this case the tendency of husbands in the disturbed group to have higher Dominance scores than husbands in the normal group is somewhat dampened when the nature of the subtest favors women's scores.
What may be the same tendency of a subtest which favors women to inhibit husband Dominance scores appears, finally, in a sex x IQ status interaction which was noted in the analysis of variance. In this case the interaction in the case of subtest 1 (Comprehension) was significant, though comparatively small (F = 5.6253, p < .05). In the case of the Similarities subtest it was again significant (F = 11.1002, p < .01) and larger, perhaps suggesting that, where the subtest favored them, the more intelligent women were more likely to achieve a high Dominance score, where on a subtest which favored the men, the tendency for one sex to dominate more than the other was strong enough to more completely muffle another tendency—that of the more intelligent partner to dominate. On subtest 3 (Vocabulary) where Dominance scores were almost equal between those couples where the wife was the more intelligent, neither the tendency of the husband to dominate nor the interaction between sex and intelligence status reached the level of statistical significance.

The trends in this analysis of variance, thus, appeared to be both internally consistent and supportive as far as they went of Bauman and Roman's findings. The fact that the nature of the subtest affects Dominance scores, however, will mean that reasons for redefining the item pool will have to include consideration for this new factor. It has already been suggested that the interaction
test form of the Wechsler Bellevue might gain in discrimi-
natory power by the addition of more items. One would
have to be aware, on the other hand, that to load this
item pool with items which would be supposed to favor one
sex over the other would probably (if the results of the
present study prove to be sound) artificially inflate or
counteract the apparent general tendency of husbands
responses to dominate over wives.

In conclusion it was pointed out that a careful
construction of a standardized item pool would possibly be
only one element of a more sophisticated methodology.
Another element might involve some slight changes in
scoring methods as well as greater clarity in scoring
criteria. These were seen as immediately necessary
attempts to refine and standardize procedures if other
researchers were to capitalize on the apparent promise of
interaction testing as a tool both of diagnosis and of
therapy. It seemed, finally that although interaction
testing is likely to offer a hitherto untapped and rich
vein of material for clinical interpretation, if it is to
prove itself as a truly psychometric device it will have to
indicate a much greater discriminatory power between normal
and other groups than it did in the present study where it
appeared to perform less adequately than a long-standing
psychometric device such as the MMPI.
APPENDIX A

Number _______  Date _______

birth date  Ordinal position (in family of origin)

1. Husband
   Wife

2. Date of marriage -
   Separations -  yes  no
   Previous marriages -  yes  no  dates

3. Number of children -
   Birth dates and sex -

   Education  Occupation

4. Husband
   Wife

   Education  Occupation

5. Husband's:  Father
   Mother

   Wife's:  Father
   Mother

6. Has there been previous treatment, therapy, or recommendation for treatment?  Yes  No

   Religion

7. Husband -
   Wife -

   Religion

8. Husband's:  Father
   Mother

   Wife's:  Father
   Mother
9. Husband's parents: living divorced separated deceased previous marriages?

10. Wife's parents: living divorced separated deceased previous marriages?

11. Referral source:
REFERENCES


Roman, M., and Bauman, G. Interaction testing: a progress report on the development of a technique for the assessment of family interaction and decision-making. Mimeo, Albert Einstein College of Medicine, Yeshiva University, 1966.


