through an oversight, the Humm-Wadsworth Temperament Scale was omitted from the list of tests included in the abstract of Irwin, R. Randall, "Lockheed's Full Testing Program," Personnel Journal, XXI (1942), 103-107, on page 89 in Volume 3, Number 1, of this journal.

AN EXPERIMENTAL ANALYSIS OF FIVE TESTS OF "APPRECIATION OF HUMOR"

H. J. EYSENCK
Psychology Department, Mill Hill Emergency Hospital, London

Introduction

It is generally agreed, among psychologists as well as among laymen, that "sense of humor" is an important and valuable personality trait. It has been equated with "insight" (2), and it has been made into a fundamental philosophical "Lebensgefühl" (23); it has been ascribed to various nations (9) and races (24) in varying proportions, usually determined by the nationality or race of the writer; it has been used as an aid in classifying and diagnosing mental illness (8, 10, 20, 29, 33, 35); it has been correlated with personality and temperament (11, 19, 25, 28, 31), as well as with scholastic aptitude, emotional maturity, height, and weight (37). Yet in spite of these manifold uses of the term, scientific measurement of the trait has lagged seriously behind.

The following methods are the only ones which appear to have been employed to date in an attempt to quantify this elusive trait; none of them can escape serious criticisms on the theoretical level, and none of them can be said to have gained wide acceptance.

(1) One common method is that of determining a person's "humor" by comparing his ranking of humorous items with that of a standard group. This is the method used, for instance, in the Roback test (34).

(2) A similar method, adopted by Almack in his test (1), and also used by other investigators (4), consists in comparing the absolute judgments of funniness of various humorous items with a standard derived from a large group.

The Experiment

Material. The experimental material consisted of five sets of twelve items each. The sets were named Jokes, Pictures, and also used by other investigators (4), consists in comparing the absolute judgments of funniness of various humorous items with a standard derived from a large group.

The Experiment

Material. The experimental material consisted of five sets of twelve items each. The sets were named Jokes, Pictures, and also used by other investigators (4), consists in comparing the absolute judgments of funniness of various humorous items with a standard derived from a large group.

The Experiment

Material. The experimental material consisted of five sets of twelve items each. The sets were named Jokes, Pictures, and also used by other investigators (4), consists in comparing the absolute judgments of funniness of various humorous items with a standard derived from a large group.
The average intercorrelations between the rankings in each of the tests are shown in Table 3. These intercorrelations, calculated from the values in Table 1 by means of Kelley's formula (27), are surprisingly small. On the average the men's rankings correlate only to the extent of .096, and the women's to the extent of .131. The average of all the 24,750 correlations is .110. In the writer's previous investigation an average correlation of .096 with the "true" order of the two sets, it was seen that the best-liked item in any test has a position of 4.2, while the least-liked item has a position of 8.9; this difference in the position of 4.7 points contrasts with a maximum possible difference of 11 points. The standard deviations show considerable variation; the least variable is a S.D. of 2.70, and the lowest variable a S.D. of 3.85. The average S.D.'s of the five tests are: Jokes = 3.18; Verses = 3.29; Comparisons = 3.17; Pictures = 3.29; Limericks = 3.17; Verses = 3.29; Comparisons = 3.17. It may be concluded that there is approximately as much divergence of opinion about the "funniness" of the items in one of the five tests as there is about that of the items in any of the others.

**TABLE 3**

<table>
<thead>
<tr>
<th>Item</th>
<th>Average InterCorrelation Between Rankings in Each Test</th>
<th>Average InterCorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jokes</td>
<td>.096</td>
<td>.067</td>
</tr>
<tr>
<td>Verses</td>
<td>.067</td>
<td>.076</td>
</tr>
<tr>
<td>Limericks</td>
<td>.121</td>
<td>.148</td>
</tr>
<tr>
<td>Comparisons</td>
<td>.121</td>
<td>.150</td>
</tr>
<tr>
<td>Mean</td>
<td>.056</td>
<td>.131</td>
</tr>
</tbody>
</table>

The average intercorrelations between the rankings in each of the tests are shown in Table 3. These intercorrelations, calculated from the values in Table 1 by means of Kelley's formula (27), are surprisingly small. On the average the men's rankings correlate only to the extent of .096, and the women's to the extent of .131. The average of all the 24,750 correlations is .110. In the writer's previous investigation an average correlation of .096 with the "true" order of the two sets, it was seen that the best-liked item in any test has a position of 4.2, while the least-liked item has a position of 8.9; this difference in the position of 4.7 points contrasts with a maximum possible difference of 11 points. The standard deviations show considerable variation; the least variable is a S.D. of 2.70, and the lowest variable a S.D. of 3.85. The average S.D.'s of the five tests are: Jokes = 3.18; Verses = 3.29; Comparisons = 3.17; Pictures = 3.29; Limericks = 3.17; Verses = 3.29; Comparisons = 3.17. It may be concluded that there is approximately as much divergence of opinion about the "funniness" of the items in one of the five tests as there is about that of the items in any of the others.

**TABLE 3**

<table>
<thead>
<tr>
<th>Item</th>
<th>Average InterCorrelation Between Rankings in Each Test</th>
<th>Average InterCorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jokes</td>
<td>.096</td>
<td>.067</td>
</tr>
<tr>
<td>Verses</td>
<td>.067</td>
<td>.076</td>
</tr>
<tr>
<td>Limericks</td>
<td>.121</td>
<td>.148</td>
</tr>
<tr>
<td>Comparisons</td>
<td>.121</td>
<td>.150</td>
</tr>
<tr>
<td>Mean</td>
<td>.056</td>
<td>.131</td>
</tr>
</tbody>
</table>

The average intercorrelations between the rankings in each of the tests are shown in Table 3. These intercorrelations, calculated from the values in Table 1 by means of Kelley's formula (27), are surprisingly small. On the average the men's rankings correlate only to the extent of .096, and the women's to the extent of .131. The average of all the 24,750 correlations is .110. In the writer's previous investigation an average correlation of .096 with the "true" order of the two sets, it was seen that the best-liked item in any test has a position of 4.2, while the least-liked item has a position of 8.9; this difference in the position of 4.7 points contrasts with a maximum possible difference of 11 points. The standard deviations show considerable variation; the least variable is a S.D. of 2.70, and the lowest variable a S.D. of 3.85. The average S.D.'s of the five tests are: Jokes = 3.18; Verses = 3.29; Comparisons = 3.17; Pictures = 3.29; Limericks = 3.17; Verses = 3.29; Comparisons = 3.17. It may be concluded that there is approximately as much divergence of opinion about the "funniness" of the items in one of the five tests as there is about that of the items in any of the others.

**TABLE 3**

<table>
<thead>
<tr>
<th>Item</th>
<th>Average InterCorrelation Between Rankings in Each Test</th>
<th>Average InterCorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jokes</td>
<td>.096</td>
<td>.067</td>
</tr>
<tr>
<td>Verses</td>
<td>.067</td>
<td>.076</td>
</tr>
<tr>
<td>Limericks</td>
<td>.121</td>
<td>.148</td>
</tr>
<tr>
<td>Comparisons</td>
<td>.121</td>
<td>.150</td>
</tr>
<tr>
<td>Mean</td>
<td>.056</td>
<td>.131</td>
</tr>
</tbody>
</table>

The average intercorrelations between the rankings in each of the tests are shown in Table 3. These intercorrelations, calculated from the values in Table 1 by means of Kelley's formula (27), are surprisingly small. On the average the men's rankings correlate only to the extent of .096, and the women's to the extent of .131. The average of all the 24,750 correlations is .110. In the writer's previous investigation an average correlation of .096 with the "true" order of the two sets, it was seen that the best-liked item in any test has a position of 4.2, while the least-liked item has a position of 8.9; this difference in the position of 4.7 points contrasts with a maximum possible difference of 11 points. The standard deviations show considerable variation; the least variable is a S.D. of 2.70, and the lowest variable a S.D. of 3.85. The average S.D.'s of the five tests are: Jokes = 3.18; Verses = 3.29; Comparisons = 3.17; Pictures = 3.29; Limericks = 3.17; Verses = 3.29; Comparisons = 3.17. It may be concluded that there is approximately as much divergence of opinion about the "funniness" of the items in one of the five tests as there is about that of the items in any of the others.

**TABLE 3**

<table>
<thead>
<tr>
<th>Item</th>
<th>Average InterCorrelation Between Rankings in Each Test</th>
<th>Average InterCorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jokes</td>
<td>.096</td>
<td>.067</td>
</tr>
<tr>
<td>Verses</td>
<td>.067</td>
<td>.076</td>
</tr>
<tr>
<td>Limericks</td>
<td>.121</td>
<td>.148</td>
</tr>
<tr>
<td>Comparisons</td>
<td>.121</td>
<td>.150</td>
</tr>
<tr>
<td>Mean</td>
<td>.056</td>
<td>.131</td>
</tr>
</tbody>
</table>

The average intercorrelations between the rankings in each of the tests are shown in Table 3. These intercorrelations, calculated from the values in Table 1 by means of Kelley's formula (27), are surprisingly small. On the average the men's rankings correlate only to the extent of .096, and the women's to the extent of .131. The average of all the 24,750 correlations is .110. In the writer's previous investigation an average correlation of .096 with the "true" order of the two sets, it was seen that the best-liked item in any test has a position of 4.2, while the least-liked item has a position of 8.9; this difference in the position of 4.7 points contrasts with a maximum possible difference of 11 points. The standard deviations show considerable variation; the least variable is a S.D. of 2.70, and the lowest variable a S.D. of 3.85. The average S.D.'s of the five tests are: Jokes = 3.18; Verses = 3.29; Comparisons = 3.17; Pictures = 3.29; Limericks = 3.17; Verses = 3.29; Comparisons = 3.17. It may be concluded that there is approximately as much divergence of opinion about the "funniness" of the items in one of the five tests as there is about that of the items in any of the others.

**TABLE 3**

<table>
<thead>
<tr>
<th>Item</th>
<th>Average InterCorrelation Between Rankings in Each Test</th>
<th>Average InterCorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jokes</td>
<td>.096</td>
<td>.067</td>
</tr>
<tr>
<td>Verses</td>
<td>.067</td>
<td>.076</td>
</tr>
<tr>
<td>Limericks</td>
<td>.121</td>
<td>.148</td>
</tr>
<tr>
<td>Comparisons</td>
<td>.121</td>
<td>.150</td>
</tr>
<tr>
<td>Mean</td>
<td>.056</td>
<td>.131</td>
</tr>
</tbody>
</table>
A slight positive (insignificant) correlation had been found in the previous experiment between liking a large number of items, and having a high score (11, p. 299). In the present experiment the following correlations were found for the five tests: Jokes = .079; Pictures = .009; Limericks = .128; Verses = .065; Comparisons = .036. The average of these correlations is only .049, which is clearly not significant.

Several interesting group factors had been found in the previous experiment (11, pp. 299-300), and an attempt was made to determine the presence and relative importance of group factors in the present series of rankings. The method used was slightly unusual, and is based on the theorem that with the exception of the first, or general, factor, the factors extracted from a matrix of correlations between persons are identical with the factors extracted from a matrix of correlations between tests (or test items).

Now, the scores derived from correlating each person's rankings with the average rankings of items are effectively approximations to centroid factor saturations derived from the matrix of intercorrelations between the subjects. In order to determine the existence and nature of group factors we can either (a) calculate the 24,750 intercorrelations of the 100 subjects for the 5 tests, factorize the resulting matrices, extracting first of all the general, positive factors to which our scores approximate, and then re-analyze the residual matrices for group factors, or (b) we can test content, with our scores as approximations to the general factor saturations and go on to correlate the items in each of the tests. In accordance with the law quoted in the preceding paragraph, this latter procedure should eliminate the general, larger factor and produce exactly the same group factors which we would have got by means of the first procedure. As the labor involved in the first method outlined above is prohibitive, the second method was used in this research.

This method, apart from being much more economical, has the added advantage that the items defining the group factors are obtained directly, and do not have to be inferred from the

1 Full discussion, formulation, and proof of this statement are given in (5).
no subtlety of any kind. Examples of clever jokes, as determined by the analysis, are F, D, J, M, L; examples of funny jokes are A, B, and C. While these jokes substantiate the above, what should be emphasized is that the analysis is partly determined by the much larger number of jokes which was found to fall into much the same pattern in the writer's earlier work (11).

The writer's test also confirms a group factor found in the earlier work, viz., a factor dividing situational from personal humor, i.e., humor based on the characteristics of the persons depicted vs. humor based on the situation, irrespective of character. Examples of the former type of humor are C, F, and M, depending on the character qualities of uptight old women, A.T.S. officers, and vacuum cleaner salesmen. Examples of the latter type are A, B, and G, in which the characters of the persons are of little importance. Unfortunately it was impossible to reproduce the cartoons so as to allow the reader to form his own judgment of the adequacy of this distinction; the descriptions given in Appendix A are hardly adequate.

The "Limericks" fall into two groups which seem to correspond with the categories found suitable for the jokes, viz., clever vs. funny. The funny limericks are those which are just absurd and nonsensical, such as C, L, E, D, B, while the clever limericks contain some twist which appeals to the intelligence. Examples of the latter type are H, F, G, A, and M. These differences will be elaborated more fully in the discussion of "sense of humor" in the following section.

Discussion

Discussion may be clarified by seeking agreement on the nature of the factors which determine our appreciation of humor. The writer has applied Burt's four-factor theory (5) to a classification of factors determining aesthetic appreciation (15), and it would appear that the results sanction a similar application in the field of humor.

Following in general the argument of this previous paper (15), we find first of all a general factor, i.e., an indication that all or a certain number of observers agree on the ranking of the items, two further factors are "unique," i.e., they are peculiar to each of the judges separately. First, individual judgments may be based on personal associations and experiences. A certain amount of experimental work has been done under this heading, both in the field of aesthetics (17) and of humor (3, 26). And also we have error factors, i.e., factors which do not remain stable even within the same person, but which change from day to day. Their relative contribution is reflected in the test reliabilities.

For such samples of the population (both of subjects and of items) as were used in this experiment and in the writer's previous work, the importance of these four types of factors appears to be roughly in the ratio of 1: 1: 6: 2.

The position is quite different when we turn to the other types of factors determined by the analysis, i.e., factors determining our appreciation of different types of humor out of which a general factor might result in a greater influence of chance elements and hence in a reduction in the correlation between the tests; yet it is difficult to see how such a reduction could lead to the type of correlation between the tests which was found.

The true position would appear to be that while the factors which determine our preferences in one test of aesthetic appreciation are closely related to, or partly identical with, those which determine our preferences in other tests of the same kind, no such similarity or identity can be found among the factors determining our preferences in different types of humorous items. This conclusion, perplexing though it may be, and unclear as its implications may appear, seems to be forced on us by the results described in the previous section. If this conclusion is accepted, it follows as an important consequence that scores on the ordinary type of test of "appreciation of humor" are valueless when regarded as giving an intimate picture of the subject's general "sense of humor." They measure nothing but the subject's reaction to the test itself.

The position is quite different when we turn to the other method of analyzing the data, viz., the correlation between the judge's ratings of amusingness and the "funniness" of the items. For each of the judges separately, over all the tests, the correlation was found to be .85. This is a higher score than that found in the writer's earlier work (11) where the correlation between the judges was found to be .68. This is in line with the general trend of our results.

In this case we can see that the "funniness" of the items is determined partly by factors which are peculiar to the individual judge, and partly by factors which are peculiar to the item itself. The factor which is peculiar to the item itself is probably the most important, and it is this factor which we are interested in when we talk about the "funniness" of an item.

The factors which determine the "funniness" of an item are probably quite complex and difficult to analyze. It is quite possible that there is no single factor which determines the "funniness" of an item, but rather a combination of several factors, each of which has a different effect on the "funniness" of the item. It is also possible that the "funniness" of an item is determined by a combination of factors which are peculiar to the individual judge, and factors which are peculiar to the item itself.

In conclusion, we can say that the "funniness" of an item is determined partly by factors which are peculiar to the individual judge, and partly by factors which are peculiar to the item itself. The factor which is peculiar to the item itself is probably the most important, and it is this factor which we are interested in when we talk about the "funniness" of an item.
ing matrices factorized, in order to discover group factors. The five tests were intercorrelated, both for scores and for "number of items liked," in order to find whether there were any factors common to all tests. The following conclusions were arrived at:

1. There was as much divergence of opinion on the "funniness" of the items in one of the tests as there was on that of the items in any of the other tests.

2. On the whole, about 35% of the items were found amusing, and the rest were known.

3. Men and women did not differ with regard to the number of items liked, but the men claimed to know on the average twice as many items as the women.

4. There was a slight tendency for items which are known to be liked better.

5. The rankings of the 100 subjects in each of the tests intercorrelated to the extent of .110 on the average.

6. The rankings of the 100 subjects in each of the tests intercorrelated to the extent of .131 and .096.

7. The validity of the average rankings was very high; it equalled .96 on the average.

8. There did not appear to be any great difference between the preferences of the men and the women, as shown by their average rankings.

9. There were no significant correlations between the scores of the 100 subjects in the five tests used.

10. There was an average intercorrelation of .37 for "number of items found amusing" among the five tests.

11. There was a slight tendency for items which are known to be liked more by men than by women to be liked more by the men and less by the women; this difference was less than .10.

12. There was no significant correlation between liking for "funny" jokes and liking for "funny" humor.

13. The two factors were present in the same tests and the same items were liked in both factors.

14. The two factors were present in the same tests and the same items were liked in both factors.

15. The two factors were present in the same tests and the same items were liked in both factors.

16. The two factors were present in the same tests and the same items were liked in both factors.

17. The two factors were present in the same tests and the same items were liked in both factors.

18. The two factors were present in the same tests and the same items were liked in both factors.

19. The two factors were present in the same tests and the same items were liked in both factors.

20. The two factors were present in the same tests and the same items were liked in both factors.

APPENDIX A

Test One: Jokes

(A) Customer: "Why does that dog hang around looking at me?"
Barber: "Well, sometimes I clip off a small piece of a customer's ear."

(B) "Why do you want Limburger cheese packed in your lunch?"
asked the grocer. "Because, papa," answered the truthful little girl, "I want to send a whole hog to school."

(C) Little Willie's mother had just seen him put his thumb to his nose and wiggle his fingers at his little playmates. "Willie," she cried, "What do you mean by doing such a thing?"
"Don't worry, mother," said Willie. "They know what I mean."

(D) There was a boy in a storm at sea, and a nervous woman passenger went to the Captain. "Captain," she asked, "are we in great danger?" "Madam," he replied, "we are in the hands of God."
"Oh!" she exclaimed, "Is it as bad as that?"

(E) A woman visitor to the London Zoo asked the keeper whether the hippopotamus was a male or a female. "Madame," replied the keeper sternly, "that is a question which should be of interest only to another hippopotamus."

(F) Associate Editor of Humorous Paper: "Let's not print any more Scotch, Jewish, Ford, or Hitler jokes." Editor: "All right. I'm tired of putting out this bawdy magazine, anyway."

(G) "You are the sunshine of my life. Your smile drives every cloud away. With you at my side, I would defy the storms of life." "What is this? A proposal or a weather report?"

(H) Hubby: "One night while you were away I heard a burglar. You should have seen me coming down the stairs three at a time." Wife: "Where was he? On the roof?"

(I) Prison Visitor: "How long are you here for?" Convict: "Thirty years." Visitor: "Ah well, here's another day easily gone."

(J) Old lady: "I wouldn't cry like that, my little man. Boy: "Cry as you please, this is my way."

(K) A man living in a village outside Paris during the Revolution met a friend fresh from the city and asked what was happening. "It's awful," was the reply; "they're cutting off heads by the hundred." "Good Heaven!" said the friend; "not heads," he cried. "Why, I'm a butcher."

(L) The speech of a young man had arrived in a crabby frame of mind, looked around and beckoned the chairman. "I would like to have a glass of water on my table, if you please," he said. "To drink to?" was the chairman's idiotic question. "Oh, no," was the sarcastic retort; "when I've been speaking half an hour I do a high dive."

208 EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

Test Two: Pictures

(A) A cattle, with wings, flying through the air. One yokel pointing it out to the other, saying: "Look, a flying fortress. That'll show 'em."

(B) Soldier with camouflage branches in his helmet, reporting to his officer. A bird has made its nest in the helmet.

(C) Old woman, looking at remains of bombed house. Sees whisky bottle inside battered old gramophone, says: "I always won-"er what is common to all tests. The following conclusions were arrived at:

1. There was as much divergence of opinion on the "funniness" of the items in one of the tests as there was on that of the items in any of the other tests.

2. On the whole, about 35% of the items were found amusing, and the rest were known.

3. Men and women did not differ with regard to the number of items liked, but the men claimed to know on the average twice as many items as the women.

4. There was a slight tendency for items which are known to be liked better.

5. The rankings of the 100 subjects in each of the tests intercorrelated to the extent of .110 on the average.

6. The rankings of the 100 subjects in each of the tests intercorrelated to the extent of .131 and .096.

7. The validity of the average rankings was very high; it equalled .96 on the average.

8. There did not appear to be any great difference between the preferences of the men and the women, as shown by their average rankings.

9. There were no significant correlations between the scores of the 100 subjects in the five tests used.

10. There was an average intercorrelation of .37 for "number of items found amusing" among the five tests.

11. There was a slight tendency for items which are known to be liked more by men than by women to be liked more by the men and less by the women; this difference was less than .10.

12. There was no significant correlation between liking for "funny" jokes and liking for "funny" humor.

13. The two factors were present in the same tests and the same items were liked in both factors.

14. The two factors were present in the same tests and the same items were liked in both factors.

15. The two factors were present in the same tests and the same items were liked in both factors.

16. The two factors were present in the same tests and the same items were liked in both factors.

17. The two factors were present in the same tests and the same items were liked in both factors.

18. The two factors were present in the same tests and the same items were liked in both factors.

19. The two factors were present in the same tests and the same items were liked in both factors.

20. The two factors were present in the same tests and the same items were liked in both factors.

APPENDIX A

Test One: Jokes

(A) Customer: "Why does that dog hang around looking at me?"
Barber: "Well, sometimes I clip off a small piece of a customer's ear."

(B) "Why do you want Limburger cheese packed in your lunch?"
asked the grocer. "Because, papa," answered the truthful little girl, "I want to send a whole hog to school."

(C) Little Willie's mother had just seen him put his thumb to his nose and wiggle his fingers at his little playmates. "Willie," she cried, "What do you mean by doing such a thing?"
"Don't worry, mother," said Willie. "They know what I mean."

(D) There was a boy in a storm at sea, and a nervous woman passenger went to the Captain. "Captain," she asked, "are we in great danger?" "Madam," he replied, "we are in the hands of God."
"Oh!" she exclaimed, "Is it as bad as that?"

(E) A woman visitor to the London Zoo asked the keeper whether the hippopotamus was a male or a female. "Madame," replied the keeper sternly, "that is a question which should be of interest only to another hippopotamus."

(F) Associate Editor of Humorous Paper: "Let's not print any more Scotch, Jewish, Ford, or Hitler jokes." Editor: "All right. I'm tired of putting out this bawdy magazine, anyway."

(G) "You are the sunshine of my life. Your smile drives every cloud away. With you at my side, I would defy the storms of life." "What is this? A proposal or a weather report?"

(H) Hubby: "One night while you were away I heard a burglar. You should have seen me coming down the stairs three at a time." Wife: "Where was he? On the roof?"

(I) Prison Visitor: "How long are you here for?" Convict: "Thirty years." Visitor: "Ah well, here's another day easily gone."

(J) Old lady: "I wouldn't cry like that, my little man. Boy: "Cry as you please, this is my way."

(K) A man living in a village outside Paris during the Revolution met a friend fresh from the city and asked what was happening. "It's awful," was the reply; "they're cutting off heads by the hundred." "Good Heaven!" said the friend; "not heads," he cried. "Why, I'm a butcher."

(L) The speech of a young man had arrived in a craby frame of mind, looked around and beckoned the chairman. "I would like to have a glass of water on my table, if you please," he said. "To drink to?" was the chairman's idiotic question. "Oh, no," was the sarcastic retort; "when I've been speaking half an hour I do a high dive."


10. Elste, R. "Über die Anwendbarkeit eines Bildserientests zur (H) (I) She was as maternal as an incubator.


REFERENCEs

all told that their interviews might be overheard. Much was done to lessen distraction by concealing the microphones in goose-neck lamps placed on the table between student and counselor. The recorder was in a distant room.

The recordings were selected on the basis of five criteria: (1) sufficient clarity of the recording to assure reasonable audibility; (2) equal sampling of early, middle, and late interviews within interview series; (3) a sampling of extent of experience on the part of the counselors ranging from one month to several years; (4) a sampling of interviewers differing in counseling philosophy; and (5) approximately equal duration of the interviews. One interview extended for 50 minutes.

For one portion of the evaluation the interviews were used in their phonographic or recorded form in a manner to be described later. For another portion of the evaluation the interviews were used in the form of typescripts of the recordings.

Selection of Judges

The selection of judges, the persons who were to use the check list in identifying the procedures in the interviews, was limited to persons who upon the basis of their training and experience were considered to be professionally competent yet who ranged from rather limited to quite extensive experience. The precaution was taken to assure that differing counseling philosophies were represented among the judges just as among the counselors. It might be well to note that a recorded interview conducted by the writer was among the nineteen studied and was rated as one of the least directive. As the writer emphasizes a less directive approach in many of his interviews, he has tried at every point to institute suitable controls.

Data-Gathering Procedures

Since the judges were unfamiliar with the check list at the outset, experience in its use had to be provided. First, a judge

The Development and Evaluation of a Measure of Counseling Interview Procedures

Part II

The Evaluation

E. H. Porter, Jr.

Miss System Council for the Oregon State Public Welfare Commission

In the first section of this report there was described the development of a check list of procedures suitable for use in identifying procedures used in counseling interviews. The development was based upon a theoretical analysis of publications on how to counsel and an empirical analysis of phonographically recorded interviews. In this section of the report there are described the steps taken in the evaluation of the check list and the outcomes of the evaluation.

Selection of Interviews

The nineteen interviews studied were selected from phonographically recorded interviews gathered by the Educational Psychology Laboratory and the Psychological Clinic at the Ohio State University during the Autumn and Winter quarters of 1940-41. The majority of the interviews were between students enrolled in a remedial study class and the counselors assigned them. The interviews, usually a half hour in duration, covered personal as well as academic problems. In accordance with the policy of the Laboratory, the students were


