when a 3-point interval was used, and the lowest correlation coefficient of .802 when the interval was 10 points. The placement of the interval also influences somewhat the size of the correlation coefficient. *K. S. Yum.*

ERRATUM

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.

(3) The third method, used by Moss in his test (30), requires the subject to select the funniest ending for a joke, several different endings being provided.

(4) A fourth method, not yet officially embodied in a standardized test comparable to the Roback, Almack, and Moss tests, was used first by Claparède (7) and Harrower (21) in their work on the psychology of the higher thought processes, and is similar to the third method. The subject is presented with a cartoon, and is required to produce a funny caption for it, or with an unfinished joke, which he is required to finish. The present writer has attempted to develop this method into a standard test in some unpublished work.

Method (3) has been shown to provide a good test of intelligence but to show little evidence of any specific "humor" factor (38); the other "production" method, while undoubtedly also highly influenced by intelligence, appears more hopeful, but is as yet in the experimental stage. This leaves us for practical purposes with the two "appreciation" methods described above.

On closer scrutiny, these two methods involve two different principles. We may lay stress on the *cognitive* agreement between the subject and the criterion group, determining the number of times his judgment regarding the goodness or badness of a joke agrees with that of the standard group, or we may lay stress rather on the *affective* score of the subject, i.e., we may enquire whether he finds very many or very few jokes funny. Method (1) is better adapted to deal with *cognitive* agreement, method (2) is better adapted to deal with *affective* scores. The virtues of both methods can be combined by having the subjects rank the items, and by asking them at the same time how many items they find amusing or funny (11).

Before we can use any of these methods, we must attempt to answer a number of questions which cannot fail to arise. We must know whether there are any sex differences, and of what kind they may be. We must know whether a score in a test constructed on the above principles is specific to that test, or whether it has any predictive power with regard to other, similar tests. (In other words, we must know whether tests

of "appreciation" of humor tend to correlate together in the same way that intelligence tests do.) Similarly, we want to know which scores, *cognitive* or *affective,* as determined by one test, have greater predictive value with regard to the whole universe of possible tests. Without an answer to these and similar questions, scores on the standard tests mentioned must remain meaningless numbers, of unknown significance.

As hardly any of these questions have been answered by the authors of the above-mentioned tests, an attempt was made to provide the required information. Some of the questions asked can best be answered by means of an experiment using a large number of humorous items, and a small number of subjects whose temperament, personality, etc., can be thoroughly investigated. An experiment of this type, employing three tests containing 100, 52, and 37 items, respectively, and using 16 subjects who also took part in a temperament test, has been reported elsewhere (11). Some of the questions asked can best be answered by means of an experiment using a smaller number of items, and a large number of subjects. An experiment of this type is reported in the pages that follow.

The Experiment

Material. The experimental material consisted of five sets of twelve items each. The sets were named Jokes, Pictures, Limericks, Verses, and Comparisons: their nature can be deduced from these names. The items comprising four of these sets are included in the present article, appendix A, and a description of the fifth set (Pictures) is also given. The items in each set were labelled A through M, the letter J being omitted, as some subjects mix it up with the letter I.

Subjects. These five tests were given to 100 subjects, 50 male and 50 female. The average age of the male group was 31.1, the average age of the female group 28.3; the average age of the whole group was 29.7. The subjects formed a rather heterogeneous sample of the population, including at one end eleven graduate and post-graduate university students, and at the other end six Civil Defence Workers from one of London's

worst slum districts. The great majority of subjects were lower middle or working class. All were British by birth.

Procedure. The subjects were asked to rank the items in each of the five tests in order of "funniness," putting the most amusing one on top and the least amusing one at the bottom. They were instructed to judge entirely by their own subjective feeling in the matter, without trying to think what the majority opinion would be in each case. They were told that there were no right or wrong answers. In addition to ranking the items in each set, they were asked to state how many of the items they found amusing, and how many items they were familiar with.

TABLE 1

The Average and Standard Deviation of the Ranks of Each Item for Each Test

Item	Jokes		Pictures		Limericks		Verses		Comparisons	
	M	σ	M	σ	M	σ	M	σ	M	σ
A	6.5	3.80	6.2	3.29	5.2	3.33	4.6	3.36	5.0	3.38
B	7.8	3.19	4.7	2.93	8.3	3.23	6.7	3.29	7.9	3.06
C	6.4	3.43	5.3	3.85	6.2	3.34	4.9	3.26	6.0	3.29
D	5.7	3.64	7.3	3.22	6.1	2.96	8.4	3.24	8.5	3.39
E	4.2	3.12	6.0	2.96	6.8	3.29	7.4	2.70	8.7	2.86
F	6.9	3.12	5.6	3.32	5.3	3.42	7.4	3.33	6.8	3.29
G	7.8	3.20	7.3	3.29	4.4	3.18	6.8	3.25	5.6	3.23
H	7.2	3.19	7.5	3.28	6.4	3.03	7.5	3.27	5.2	2.72
I	5.9	2.93	8.9	2.83	8.1	2.77	6.2	3.32	5.3	3.29
K	6.3	3.28	5.4	3.06	8.1	3.30	6.3	3.26	8.0	2.96
L	7.2	3.31	6.9	3.39	8.2	3.18	5.4	3.66	5.0	3.32
M	6.2	3.53	6.9	3.50	4.9	3.03	6.5	3.51	6.2	3.37

Results. In Table 1 are given the average positions and the standard deviations of each of the twelve items in each of the five tests. It will be 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 position of 4.7 points contrasts with a maximum possible difference of 11 points. The standard deviations show considerable variation; the least variable item has a S.D. of 2.70, and the most variable item a S.D. of 3.85. The average S.D.'s of the five tests are: Jokes = 3.31; Pictures = 3.24; Limericks = 3.17; Verses = 3.29; Comparisons = 3.18. 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 2

Average Numbers of Items Liked and Known

Test	Average number liked				Average number known			
	Men	Women	Total		Men	Women	Total	
	M	M	M	σ	M	M	M	σ
Jokes	3.98	4.04	4.01	2.89	1.46	.76	1.11	1.69
Pictures	5.67	5.20	5.43	2.88	3.78	1.98	2.88	4.12
Limericks	3.90	4.00	3.95	2.72	1.70	1.51	1.61	2.90
Verses	3.36	3.86	3.61	2.94	.42	.22	.32	.94
Comparisons	4.06	4.36	4.21	2.65	.82	.46	.64	1.84
Average	4.19	4.29	4.24		1.64	.99	1.31	

In Table 2 are given the average numbers of items liked ("found amusing") and known in each of the five tests. On the whole, about 35% of the items were found amusing, and about 10% were known. While the number of items liked does not differ much from test to test (particularly considering the large S.D.'s), there are great differences among the tests regarding the number of items with which the subjects were familiar. The "Pictures" were best known, the Limericks and the Jokes were much less well known, and the Comparisons and the Verses were hardly known at all. While there was hardly any difference in the number of items liked between the men and the women (4.19 and 4.29, respectively), there was a great difference in the number of items known between men and women. The men, on the average, claimed to know almost twice as many items as the women (1.64 and .99 are the respective figures).

Items which are known are liked slightly better than items not known. On the average, the known jokes rank 5.9, the known pictures 6.2, the known limericks 5.8, the known verses 6.0, and the known comparisons 5.6, as compared with a chance average of 6.5. Whether this is due to the fact that people tend to prefer familiar material, or whether better material is more likely to get known, it is impossible to say. It might even be possible that better-liked material is better remembered.

When the number of items liked in each of the tests is plotted, the resulting distribution is normal with the exception of the values for 0 (i.e., no items liked at all), which is too

large in every single test. Summing the distributions for the five tests, we get Figure 1, which clearly illustrates this phenomenon.

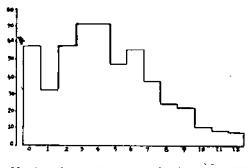

FIG. 1. Number of items liked: totals for 5 tests and 100 subjects.

TABLE 3

Average Intercorrelation Between Rankings in Each Test

Test	Average intercorrelations		
	Men	Women	Total
Jokes	.068	.067	.067
Pictures	.063	.148	.098
Limericks	.121	.174	.148
Verses	.106	.103	.090
Comparisons	.121	.165	.145
Mean	.096	.131	.110

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 .16 was found for the rankings of 187 items, grouped in three tests, by 16 subjects. The reduction in the size of the average intercorrelation from the previous work may be accounted for by the greater heterogeneity of the population;[1]

[1] When *tests* are correlated, homogeneous populations will give smaller average intercorrelations than heterogeneous populations; when *persons* are correlated, this rule is reversed. Similarly, the more homogeneous the experimental material, the

another possible explanation, greater homogeneity of the experimental material, seems unlikely as a causal factor because the same principles of selection were employed in both cases.

As the writer has shown elsewhere (12, 13), it is possible to find from the average intercorrelations of a number of rankings the correlation of the sum of these rankings with the "true" order of merit of the items ranked, i.e., with that order which would have resulted if an infinite number of subjects had been used. These values are given in Table 4; it will be seen that

TABLE 4

Correlations of Average Rankings with "True" Rankings

Test	Correlations of average with "true"		
	Men	Women	Total
Jokes	.89	.89	.94
Pictures	.87	.94	.96
Limericks	.93	.95	.97
Verses	.93	.92	.95
Comparisons	.93	.95	.97
Mean	.91	.93	.96

the average rankings of the items in all the five tests used are highly valid; on the average, our empirical rankings correlate to the extent of .96 with the "true" rankings.

If there are no differences in the factors which make men and women prefer one item to another in the tests used, then the correlations between the average rankings of the men and the average rankings of the women should equal the product of the respective correlations with the "true" order of the two groups. These theoretical values are set out below, in Table 5,

TABLE 5

Theoretical and Actual Correlations Between Rankings of 50 Men and 50 Women

Test	Theoretical	Actual
Jokes	.79	.76
Pictures	.82	.78
Limericks	.88	.91
Verses	.86	.61
Comparisons	.88	.74
Mean	.83	.76

larger will the correlations be when *tests* are correlated, while the correlations will be smaller when *persons* are being correlated.

together with the actual correlations. It will be seen that the average theoretical correlation between men's and women's rankings is .83, as compared with an actual correlation of .76. These values agree too closely to allow us to assume any considerable difference between the appreciation of humor of men and women, respectively. This conclusion is in agreement with earlier work (11, 22).

Scores for tests of the kind employed here can be derived by correlating each person's rankings in each of the tests with the average ranking of the items in those tests by the whole group (14). We have already seen that the average rankings are highly valid, correlating to the extent of .96 with the "true" order. The distribution of the resulting 500 scores, grouped about their common average in terms of sigma units, is shown in Figure 2. It will be seen that the scores are distributed roughly in the shape of the normal distribution curve.

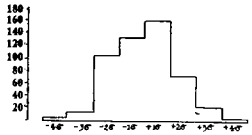

FIG. 2. Distribution of 500 scores on Appreciation of Humor tests, plotted in terms of their own sigma units.

The scores of the 100 subjects in each of the five tests can be correlated, in order to show whether or not those who score highly in one test will also tend to score highly on the other tests. In a previous paper the writer showed that the correlations between three tests of appreciation of humor were not statistically significant; similarly Stump did not find any significant correlation between the two parts of the Almack humor test (37). The correlations among the five tests employed in this study are given in Table 6; it will be seen tha

TABLE 6

Correlations Among the Five Tests

Test	Jokes	Pictures	Limericks	Verses	Comparisons
Jokes	(–)	–.03	.15	–.09	.05
Pictures		(–)	–.07	–.03	.13
Limericks			(–)	.10	.18
Verses				(–)	–.02
Comparisons					(–)
Average	.02	.00	.09	–.01	.09

there is no evidence of any tendency for those who score highly in one test to score highly in any other test. The average intercorrelation is only .0426 (the original correlations were calculated to four decimals, but are given to only two decimals in the table). This result is very different from the results of similar experiments carried out in the field of aesthetics, where the writer showed that eighteen tests of aesthetic appreciation correlated together significantly, giving rise to a general factor (called "T") which accounted for 21% of the variance (14).

In a previous experiment, it had been found that people who report that they find a large number of items amusing in one test also tend to report that they find a large number of items amusing in the other tests (11). The correlations among the five tests for number of items found amusing are given in Table 7; they average .57, as compared with an average inter-

TABLE 7

Correlations of Numbers of Items Found Amusing for the Five Tests

Test	Jokes	Pictures	Limericks	Verses	Comparisons
Jokes	(–)	.57	.67	.62	.63
Pictures		(–)	.46	.42	.47
Limericks			(–)	.65	.57
Verses				(–)	.64
Comparisons					(–)
Average	.62	.48	.60	.58	.58

correlation of .55 in the previous experiment. In both sets of experiments the "Jokes" test is the most diagnostic of this tendency, although in the one case it consisted of 100 items, in the other of only 12.

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).*[2]

Now, the scores derived from correlating each person's ranking with the average ranking are in effect 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 rest 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 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

[2] Full discussion, formulation, and proof of this statement are given in (5).

judgments of the subjects obtaining particularly high positive or negative saturations, as is the case in method (a). The advantage of this procedure will be realized by all who have had to decide just how many subjects should be used in order to delineate group factors, and which subjects' factor saturations should be deemed too low for them to be included in this determination.

First of all, the raw rankings were inspected carefully to discover those tests in which there might be a reasonable hope of finding significant and meaningful group factors. The "Jokes," "Pictures," and "Limericks" tests were considered the most likely to show evidence of group factors, and accordingly were analyzed by means of the method outlined above.

The resulting three twelve-square matrices contained positive and negative correlations in roughly equal proportions. When $N = 100$, a correlation of .20 is significant according to Fisher's method ($p = .05$), while a correlation of .25 is very significant ($p = .01$). The three matrices contained, respectively, 10, 12, and 12 significant correlations, and 4, 8, and 9 very significant correlations. (By chance, we might have expected 3 significant and $\frac{1}{2}$ very significant correlations in each matrix.) We may conclude that all three matrices show evidence of group factors.

Factor analysis of these three matrices results in three group factors which account, respectively, for 8%, 10%, and 11% of the variance. Psychologically, these factors confirm analyses made in an earlier experiment, in which a much smaller number of subjects, a much larger number of items, and a different method of factor analysis had been used. This agreement provides welcome confirmation of the validity both of the psychological analysis and of the statistical procedure here advocated.

As regards the analysis of the "Jokes," the items seem to fall into two groups which may roughly be identified as "clever" and "funny." The clever jokes are less obvious, usually circumlocutory, and generally require the mind to jump from what is said to what is implied. The funny jokes are very obvious, deal frequently with topics pertaining to the body, and demand

no subtlety of any kind. Examples of clever jokes, as determined by the analysis, are F, D, I, M, L; examples of funny jokes are A, B, and C. While these jokes substantiate the above analysis, it should be mentioned 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 "Pictures" 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 characters 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 tippling 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 the 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 there is some kind of agreement among *all* our subjects on the

"funniness" or otherwise of the jokes or other humorous items. When this general factor is eliminated by a suitable experimental or statistical procedure, *group* factors emerge, i.e., factors which characterize the preference judgments of groups of subjects. Such factors are, for instance, the "funny" vs. "clever" factor, which bears a resemblance to the "simple" vs. "complex" group factor which was found in the author's work in aesthetics. This factor appears to be correlated with temperament, the introverted subjects tending to prefer the "clever" and "complex" items, the extraverted subjects preferring the "funny" and "simple" items (16, 11).

While these two factors are "communal" in the sense 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 vary from day to day. Their relative contribution is reflected in the retest 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. In other words, while individual factors and error factors accounted for some 80% of the variance, general and group factors accounted for only about 20%. In the author's work on preference judgments in aesthetics, "communal" factors were considerably more important (15).

It is perhaps because the general factor is so much less important in the present series of tests than it was in the series of 18 tests of aesthetic appreciation used by the author in attempting to establish the general factor of aesthetic appreciation that we find no correlation between the scores in the five individual tests used. While this is a possible reason, it does not appear a very likely one; a reduction in the importance of the 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 total abolition of any correlation between the tests.

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 appreciation of 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 be 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 estimate 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 scoring, i.e., in terms of number of items found amusing. Here we find that it is possible to predicate a general factor embracing all the tests. This general factor, indeed, appears to be even stronger than that found in analyzing the correlations between the eighteen tests of aesthetic appreciation (14). This would suggest that the essence of "sense of humor" is affective rather than cognitive; or rather, that there is great individual constancy in the manner in which subjects react affectively to humorous items of any kind, but that their cognitive reactions are determined largely by factors of individual experience, and do not allow prediction from one test to another. The relevance of this finding to test construction is too obvious to be stressed. It should also be noted that our conclusion agrees well with clinical findings which have not always been taken into consideration by psychological test-constructors (6, 18, 32, 36, 39).

The results as reported above seem to strengthen the theoretical analysis of "sense of humor" given by the writer in a previous paper (11). There it was shown that three factors

demand to be taken into account in explaining the amusement derived from a joke (taking joke as the generic name for all types of humorous items). These three factors are cognitive, conative, and affective, and the three types of reaction to which they give rise were called the *Comic, Wit,* and *Humor,* respectively. An analysis of the cognitive factors showed that there are five points which enter into the joke; we may summarize them by saying that we consider something as "funny" or "amusing" when there is *a sudden, insightful integration of contradictory or incongruous ideas, attitudes, or sentiments which are experienced objectively* '(11, p. 307).[3] Affective factors were found to include a state of joyfulness, while conative factors included a feeling of "superior adaptation," as Ludovici calls it.

Correlations between group factors and the results of a temperamental analysis seemed to show that preferences for "clever" jokes indicated a prevalence of cognitive factors, and an introverted temperament, while preference for "funny" jokes indicated a prevalence of conative and affective factors, and an extraverted temperament (11).[4] The results of the present study, as far as they go, are in full agreement with this analysis.

Summary and Conclusions

Five tests of "appreciation of humor," each consisting of 12 items to be ranked in order of "funniness," were given to 50 male and 50 female subjects, representing a sample of the population fairly representative as regards social class, age, etc. In addition, the number of items known and the number of items found funny were noted for each test and for each subject. Scores were established for each person in each test by correlating the subjects' rankings with the average ranking. The items in each of three tests were intercorrelated, and the result-

[3] According to the "Law of Reinstatement by Partial Content," any one member of a causal series may in time suffice to produce an effect which was at first produced only by the action of all the members of that series. This qualification of the above principle should be kept in mind when analyzing actual instances of laughter and amusement.

[4] These results confirmed Kambouropoulou's finding that "the more extraverted subjects have a greater proportion of the superiority class among the items they find most amusing. Extraversion and preference for the superiority class of humorous items go together" (26, p. 55).

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 10% 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 women intercorrelated on the average higher than those of the men, the respective correlations being .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 .57 for "number of items found amusing" among the five tests.

(11) There was no significant correlation between liking a large number of items and having a high score.

(12) Three group factors were found in an analysis of the intercorrelations of the items in three tests:

a) "Clever" jokes vs. "funny" jokes. ("Jokes" test.)
b) "Situational" vs. "character" humor ("Pictures" test.)
c) "Clever" limericks vs. "funny" limericks. ("Limericks" test.)

(13) The results of this experiment, as far as they go, confirmed the theoretical analysis of "sense of humor" given elsewhere by the writer.

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 son, "I want teacher to send me home."

(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 heavy 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 lousy 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." Wifie: "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 nearly gone."

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

(L) 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 thousand." "Good Heavens! Surely not heads," he cried. "Why, I'm a hatter."

(M) The speaker, who 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?" 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."

Test Two: Pictures

(A) A castle, 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 branches.

(C) Old woman, looking at remains of bombed house. Sees whisky bottle inside battered old gramophone, says: "I always wondered where she kept it."

(D) Oxford don, writing on the college wall with a piece of chalk: "Septimus Perks M.A. is dementia praecox."

(E) Policeman, pointing out a sign "Diversion" to a cyclist who was riding past it: "What the devil d'you think we put that there for? A diversion or something?"

(F) Witch has turned vacuum cleaner salesman into a frog, saying: "And if the cleaner does all you say, I will remove the spell."

(G) Maid, holds the head of a dog in her hand while the headless body stands behind her: "He's been barking his head off."

(H) Goering offers Hitler some alcohol at a party. "Well, just the teeniest bit, as it's Christmas—but I mustn't get into wrong ways."

(I) Ghost holds his head in amazement as owner of castle walks along the corridor with his head under his arm.

(K) Hitch-hiker signalling bomb-disposal car with live bomb in it.

(L) Air-raid wardens in front of shelter, talking about a third warden: "Weird sort of cove: each time a bomb drops he says 'that's 'andy.'"

(M) Contrite A.T.S. officer being addressed by superior officer in front of other A.T.S.: "For conduct unbecoming to an officer and a gentlewoman, Captain Winkworth will be 'kept in' for a week."

Test Three: Limericks

(A) There was an old party of Lyme,
Who married three wives at one time,
When asked, "Why the third?"
He replied, "One's absurd,
And bigamy, sir, is a crime."

(B) There was a young man of Laconia,
Whose mother-in-law had pneumonia;
He hoped for the worst,
And after March first
They buried her 'neath a begonia.

(C) There was a young girl of Asturias,
Whose temper was frantic and furious.
She used to throw eggs
At her grandmother's legs
A habit unpleasant, but curious.

(D) There was an old man of Tralee
Who was bothered to death by a flea;
So he put out the light,
Saying, "Now he can't bite,
For he'll never be able to see."

(E) There was a young lady of Ealing,
Who had a peculiar feeling
That she was a fly,
And wanted to try
To walk upside down on the ceiling.

(F) There was a young woman named Bright,
Whose speed was much faster than light,
She set out one day,
In a relative way,
And returned on the previous night.

(G) There was a young lady named Starkey,
Who foolishly married a darkey.
And then for her sins,
She had three pairs of twins,
One white, one black, and one khaki.

(H) For beauty I am not a star,
There are others more handsome by far;
But my face I don't mind it,
For I am behind it,
It's the people in front that I jar.

(I) There's a very mean man of Belsize,
Who thinks he is clever and wise.
And what do you think?
He saves gallons of ink
By simply not dotting his "i's."

(K) A prideful young lady of Boston
A two-horned dilemma was tossed on;
As to which was the best,
To be rich in the West
Or poor and peculiar in Boston.

(L) There was an old soldier of Bister
Went walking one day with his sister;
When a cow at one poke
Tossed her into an oak
Before the old Gentleman missed her.

(M) An epicure, dining at Crewe,
Found quite a large mouse in his stew
Said the waiter, "Don't shout,
And wave it about,
Or the rest will be wanting one, too."

Test Four: Verses

(A) Here lies the body of poor old Jones,
Who all his life collected bones.
Until came death, that mighty spectre,
That all-amassing bone collector,
And boned old Jones so neat and tidy,
That here he lies, all *bona fide*.

(B) Lady, lady, should you meet
One whose ways are all discreet,
One who murmurs that his wife
Is the lodestar of his life,
One who keeps assuring you
That he never was untrue,
Never loved another one—
Lady, lady, better run.

(C) How courteous is the Japanese,
He always says, "Excuse me please."
He climbs into his neighbour's garden
And smiles and says, "I beg your pardon."
He bows and grins a friendly grin,
And calls his hungry family in;
He grins and bows a friendly bow:
"So sorry, this *my* garden now."

(D) Or ever a lick of Art was done,
Or ever a one to care,
I was a Purple Polygon,
And you were a Sky-Blue Square.
I yearn for you, but I have no chance,
You lie in a different plane,
I break my heart for a single glance,
But I break said heart in vain.

(E) The Zebra, born both black and white,
Is just the jungle clown,
The lionesses hunt him up,
The lions hunt him down.
His life, in consequence is brief
And seems inclined to end in grief.
And so, you see, between the two,
He's more contented in the zoo.

(F) This cheerful story tells the fate
Deserved by William Smith—the late—
When he had killed two lions, he
Was killed in turn, by number three.
The lion's notable behavior
Was printed in the roto-gravure.

(G) The crocus grows in any spot,
And multiplies an awful lot.

It doesn't pout, and fail to bloom,
Because of soil, or lack of room.
No books are written on the crocus;
It grows without such hocus-pocus.

(H) Marco Polo travelled far,
Went through perils worse than war,
Saw great marvels, distant lands,
Unicorns and sarabands.
Why, and wherefore? So that when
He had toiled back home again,
He could chant an endless solo
On the deeds of Marco Polo.

(I) Down in the silent hallway
Scampers the dog about,
And whines, and barks, and scratches,
In order to get out.
Once in the glittering starlight,
He straightway doth begin
To set up a doleful howling
In order to get in.

(K) The Horse is not supposed to know
How to reap or how to sow;
How to boast or how to bind
Dictionaries in the Mind;
How to build a rabbit-hutch—
But it doesn't matter much,
For he understands, of course,
Exactly how to be a horse.

(L) Gorillas, with intent to please
The simple-minded, bend their knees
And walk through equatorial lands
Supported on their horny hands.
—A practice, it is fair to add,
That many now admit is bad
And only have recourse to in
Emergency, or after gin.

(M) I recollect a nurse call'd Ann,
Who carried me about the grass,
And one fine day a fine young man
Came up, and kiss'd the pretty lass.
She did not make the least objection;
Thinks I, "Aha."
"When I can talk I'll tell Mamma."
—And that's my earliest recollection.

Test Five: Comparisons

(A) He came into the room like a squirt from a syphon.
(B) He took in the room with a glance like a lasso.

(C) He adjusted his face as though it were a necktie.
(D) A croquette is nothing but a hash that has come to a head.
(E) My heart was going like a woodpecker at my side.
(F) He had no more initiative than a shadow.
(G) She approached with the slow dignity of a ferry coming into dock.
(H) He was as inconsiderate as an alarm clock.
(I) She was as maternal as an incubator.
(K) He was as irresponsible as a streak of lightning.
(L) She turned him down like a bedspread.
(M) The hours crawled by like paralytic centipedes.

REFERENCES

1. Almack, J. C. *Sense of Humor Test* (Form I). Cincinnati: C. A. Gregory, 1928.
2. Allport, G. W. *Personality.* London: Constable, 1938.
3. Barry, H. "The Role of Subject-Matter in Individual Differences in Humor." *Journal of Genetic Psychology,* XXXV (1928), 112–128.
4. Bird, G. "An Objective Humor Test for Children." *Psychological Bulletin,* XXII (1925), 137–138.
5. Burt, C. *The Factors of the Mind.* London: University of London Press, 1940.
6. Casella, P. *Contribution a l'Etude du Rire et du Pleurer Spasmodiques.* Lyons, 1902.
7. Claparede, E. "La Genèse de l'Hypothése." *Archives de Psychologie,* XXIV (1934), 1–154.
8. Coriat, I. H. "Humor and Hypomania." *Psychiatric Quarterly,* XIII (1939), 681–688.
9. Egner, F. "Humor und Witz unter strukturpsychologischem Gesichtspunkt." *Arch. f. d. ges. Psychol.,* LXXXIV (1932), 330–372.
10. Elste, R. "Über die Anwendbarkeit eines Bildserientests zur Beurteilung psychischen Krankheitsgeschehens." *Ztsch. ges. Neurol. Psychiat.,* CLXX (1940), No. 1.
11. Eysenck, H. J. "The Appreciation of Humour: An Experimental and Theoretical Study." *British Journal of Psychology,* XXXII (1942), 295–309.
12. ————. "The Validity of Judgments as a Function of the Number of Judges." *Journal of Experimental Psychology,* XXV (1939), 650–654.
13. ————. "The Validity and Reliability of Group Judgments." *Journal of Experimental Psychology,* XXIX (1941), 427–434.
14. ————. "The General Factor in Aesthetic Judgments." *British Journal of Psychology,* XXXI (1940), 94–102.
15. ————. "The Experimental Study of the 'Good Gestalt'—A New Approach." *Psychological Review,* XLIX (1942), 344–364.

16. ————. "Abnormal Preference Judgments as 'Complex' Indicators." *American Journal of Orthopsychiatry,* XII (1942), 338–345.
17. ————. "Some Factors in the Appreciation of Poetry, and Their Relation to Temperamental Qualities." *Character and Personality,* XI (1940), 160–167.
18. Fyodoroff, M. I. (On irresistible laughter in some forms of nervous affections.) Vestnik klin. i subebuoi psychiat. i nevropatol., VIII (1890–91), 78–97. Abstracted in: *Ann. de psychiat. et d'hypnol.,* II (1892), 87.
19. Gregg, A., Miller, M. Lindon, E. "Laughter Situations as an Indication of Responsiveness in Young Children." In: *Some New Techniques for Studying Social Behavior.* New York: D. S. Thomas, 1929.
20. Haggard, E. A. "A Projective Technique Using Comic Strip Characters." *Character & Personality,* X (1942), 289–295.
21. Harrower, M. "Organization of Higher Mental Processes." *Psychol. Forsch.,* XVII (1932), 56–120.
22. Heim, A. "An Experiment on Humour." *British Journal of Psychology,* XXVII (1936), 148–161.
23. Höffding, H. *Humor als Lebensgefühl.* Leipzig: Teubner, 1918.
24. Kadner, S. *Rasse und Humor.* Munich: Lehmann, 1939.
25. Kambouropoulou, P. "Individual Differences in the Sense of Humor." American Journal of Psychology, XXXVII (1926), 288–297.
26. ————. "Individual Differences in the Sense of Humor and Their Relation to Temperamental Differences." *Archives of Psychology,* 1930, No. 121.
27. Kelley, T. L. *Statistical Methods.* New York: Macmillan, 1924.
28. Landis, C. and Ross, J. "Humor and its Relation to Other Personality Traits." Journal of Social Psychology, IV (1933), 156–175.
29. Mayer-Gross, W. "Über Spiel, Scherz, Ironie und Humor in der Schizophrenie." *Ztsch. f. d. ges. Neurol. und Psychiat.,* LXIX (1921), 332–353.
30. Moss, F. A., et al. *George Washington Test of Social Intelligence.* 1926.
31. Omwake, L. "A Study of Sense of Humour: Its Relation to Sex, Age, and Personal Characteristics." Journal of Applied Psychology, XXI (1937), 688–704.
32. Oppenheim, H. "Über Lachschlag" (Lach-schwindel; Lach-Ohnmacht.). *Monatsch. f. Psychiat. und Neurol.,* XI (1902), 241–247.
33. Rebatel, A. *Le Rire dans la Démence Précoce.* Lyons: 1908.
34. Roback, A. A. *Sense of Humor Test.* Cambridge, Mass., 1939.
35. Senise, T. *Il Riso in Fisio-Patologia.* Naples: Bibl. d. Stud. 1941.

36. Sikorski, I. A. (Uncontrollable laughter and crying.) *Protok. Kiev Prikhsat. Obst.*, II (1899–1900), 13–20.
37. Stump, N. "Sense of Humor and its Relationship to Personality, Scholastic Aptitude, Emotional Maturity, Height and Weight." Journal of General Psychology, XX (1939), 25–32.
38. Thorndike, R. L. and Stein, S. "An Evaluation of the Attempts to Measure Social Intelligence." Psychological Bulletin, XXXIV (1937), 275–285.
39. Wiesner, A. (The pathology of laughter.) *Casop. lék. česk.*, 42, 1903, 141, 161, 201, 226, 251.

THE DEVELOPMENT AND EVALUATION OF A MEASURE OF COUNSELING INTERVIEW PROCEDURES

Part II

The Evaluation

E. H. PORTER, JR.
Merit System Council for the Oregon State Public Welfare Commission[1]

In the first section[2] 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.[3] The majority of the interviews were between students enrolled in a remedial study course 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

[1] On military leave of absence.
[2] Porter, E. H., Jr., The Development and Evaluation of a Measure of Counseling Interview Procedures: Part I, The Development. *Educational and Psychological Measurement*, III (1943), 105–126.
[3] For a complete discussion of the recording program, the apparatus used, difficulties encountered, and transcription methods, see Covner (1941). (Bibliography contained in first section of this report.)

all told that their interviews might be overheard. Much was done to lessen distraction by concealing the microphones in gooseneck 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:[4] (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 would be acknowledged 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.[5] 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

[4] Complete information regarding the recordings selected may be obtained from the writer's dissertation bearing the same title as these reports. The dissertation is on file with the library at Ohio State University.
[5] In the first section of this report directiveness was defined as the extent to which the counselor imposes upon the interview the direction which it takes.

was given written instructions describing the general outline of the activities in which he was to engage, written illustrations of each check list category, and a sample interview in typewritten form with the counselor responses identified on the margins of the pages. Following the written instructional material each judge worked with two typescripts, coding on the left margins each identified response. The judge and writer then compared their codings and attempted to clarify the meaning of each category. After two additional typescripts the judge and writer again conferred. At the conclusion of each typescript the judge was to indicate on an eleven-point scale his judgment of the directiveness of the interviewing counselor without going back over the interview in any way.

When the judge had completed the preliminary training he used the check list on four phonographically recorded interviews. As he listened he classified each identified procedure of the counselor by entering a tally in the appropriate space before the category on a mimeographed copy of the check list.[6] When a recording had been completed by a judge he indicated his evaluation of the directiveness of the interviewing counselor by encircling the appropriate number in the scale given at the bottom of the check list. There were no particular controls on the auditory conditions except to assure reasonably adequate perception of the speech.

Both the typescripts and recordings were so randomized in presentation that each judge worked with interviews that had been conducted by counselors of differing viewpoint and with interviews from the early, middle, and late periods within interview series. Although each interview was worked on by two judges in addition to the writer, no two judges worked with more than one interview in common except in the case of the first typescript and first recording, which all judges did alike. The writer took the precaution of spacing his work with the two forms of each interview no less than a week apart. Meanwhile, of course, work on other interviews had been interposed.

[6] The check list and category codings are reproduced in this section for the reader's convenience.