Published as a separate and in *The Journal of Psychology*, 1941, **12**, 111-124.

A NEW MEASURE OF INTROVERSION-EXTROVERSION*

Department of Educational Psychology, University of Minnesota

CATHARINE EVANS AND T. R. MCCONNELL

A. INTRODUCTION

This paper describes the development of relatively independent measures for three types of Introversion-Extroversion,[1] Thinking, Social, and Emotional. The need for clarifying the concept of *I-E* and for devising new inventories can best be understood by reviewing the confusion concerning its nature and measurement. In the effort to simplify the original complex description of *I-E* by Jung, psychologists either have introduced new concepts or emphasized varying phases of Jung's definition. In this process of elaboration, they have actually complicated rather than clarified the idea of *I-E*. The use of these terms in the popular literature has only added to the confusion. Unfortunately, introversion, at least in the popular writings on psychology, has come to denote an undesirable personality tendency which borders on a neurotic condition.

In general, the available *I-E* inventories purport to measure a general, undifferentiated trait. However, the intercorrelations between the published inventories are surprisingly low. Only five of the 19 coefficients of intercorrelation reported in the literature for nine inventories are above .40, and only two are above .80. The two coefficients above .80 are between two inventories and revised forms of these same inventories.

A plausible explanation for the low intercorrelations of the *I-E* inventories is the variation in the definitions on which they are based. The authors of these inventories have emphasized varying phases or types of *I-E* in their definitions of the general trait. For example, Laird and Marston have been concerned with the differences in the emotional and affective reactions of the introvert and extrovert. In contrast, Conklin has stressed the thinking reactions

*Received in the Editorial Office on June 30, 1941, and published immediately at Provincetown, Massachusetts.

[1]For the remainder of the article, Introversion-Extroversion will be designated as *I-E*.

of the introvert and extrovert; he has dealt with the more intellectual interests of the introvert as contrasted to the greater interest of the extrovert in overt activity. The differentiation of social reactions introduced by Freyd in a summary list of 54 characteristics of the introvert can be observed in items of the inventories of Bernreuter, Heidbreder, and others. Several authors have emphasized the extreme forms of *I-E* behavior. For example, the research of Neymann-Kohlstedt and Morgan-Gilliland has been dominated by the conception of manic depressive insanity and hysteria as extreme extrovertive phenomena and of schizophrenia, psychasthenia, and neurasthenia as extreme introvertive behavior. Although the *I-E* inventories published before 1940 were supposedly constructed to measure a general trait, it seems necessary to conclude that they measure different aspects of *I-E*. The conceptions underlying them vary greatly, and the intercorrelations are sufficiently low to indicate that they are not measuring the same trait.

The available evidence on the validity of the published *I-E* inventories yields additional indication that these tests are not measuring the same trait. The results of the research on the differentiation of age groups, of the sexes, and of groups varying in educational interests are not consistent from test to test and even for the same test with different groups of individuals. It is also true that none of the inventories published before 1940 has consistently displayed a degree of reliability sufficiently high for individual prediction. In fact, only eight of the 29 coefficients of reliability found in the literature for nine *I-E* inventories are above .85, and the lack of information on the number of cases makes the significance of four of these eight coefficients questionable.

In the educational and vocational guidance of students, the employment of these *I-E* inventories has brought bewilderment. Introversion as measured by Conklin is related to scholastic success, while introversion as measured by Bernreuter is correlated highly with neurotic tendency. No general conclusion as to the usefulness of the *I-E* inventories has been possible, for it is a function of each inventory.

Guilford (1) has made a valuable contribution to the measurement of *I-E*. His research has yielded conclusive evidence that the items in the available *I-E* tests are not measuring a single dimension of personality. Factor analyses of 35 typical *I-E* items revealed

several independent factors, such as Thinking *I-E,* Social *I-E,* Emotional *I-E,* Masculinity-Femininity, and Freedom from Care.

Guilford has also attacked the problem of constructing *I-E* inventories to measure specific factors rather than a general, undifferentiated trait. In 1940 he published *An Inventory of Factors STDCR* which yields measures for the following five dimensions: Social Introversion, Thinking Introversion, Depression, Cycloid Tendencies, and Rhathymia or Freedom from Care. This inventory was developed on the basis of factor analyses of the 35 typical *I-E* items (1) and of the factor analysis of 30 items constructed to emphasize the Thinking and Rhathymia factors (2).

Guilford seems to have included in the preliminary form of *An Inventory of Factors STDCR* many items in addition to the ones involved in his factor analyses. In fact, the published form is composed of 175 items retained after successive tests of internal consistency. A majority of the items in the inventory are scored for more than one factor, and some items are scored for as many as four of the five factors with the factor weights varying from one to two points.

As yet no published data on the validity of *An Inventory of Factors STDCR* seem to be available with the exception of the factor analysis studies. However, the reliability of the factor scorings has been reported (3). The estimates of the reliability of the scoring for the five factors vary from .84 to .94. The inventory thus seems sufficiently reliable for individual prediction.

Guilford, like Bernreuter, has been unsuccessful in obtaining noncorrelated tests with multiple scoring of items. The intercorrelation coefficients reported by Guilford (3) vary from —.02 to .85 with five of the 10 coefficients equal to or above .33. These five coefficients are as follows: between *D* and *S,* .49; between *C* and *S,* .33; between *C* and *D,* .85; between *R* and *S,* —.54; and between *R* and *D,* —.36. Although his early analysis of the 35 typical *I-E* items had revealed relatively independent factors, substantial intercorrelations among the factor scorings of the published inventory occur. Guilford, therefore, has not yet devised relatively independent measures of the five types of *I-E.*

The purpose of the study reported here was to develop relatively independent measures for three *I-E* factors or types, Thinking, Social, and Emotional. These factors were clearly differentiated by

Guilford (1) and it was thought that their measurement might ultimately have significance for personality diagnosis and guidance. The problem, therefore, was to devise three discrete, or practically discrete measures. This Guilford, by scoring the same items for more than one trait, has not accomplished.

B. Construction of the *I-E* Inventory

The effort to develop three homogeneous and practically independent measures determined the method of construction of the *I-E* inventory. First of all, the three *I-E* types were carefully and precisely defined. In formulating the definitions for this investigation, the introvert and extrovert of each type were differentiated in terms of two criteria suggested by Jung (4): first, the introvert is more oriented to or governed by subjective factors while the extrovert is more oriented to or governed by objective and external conditions; second, the direction of the response of the introvert tends inward, but the direction of the response of the extrovert is outward toward the object. The writings of Guilford, Conklin, McDougall, and Freyd, as well as Jung, have influenced the formulation of the definitions employed in this study. Brief statements of the definitions of the three types of *I-E* which were finally adopted follow:

The thinking introvert shows a liking for reflective thought, particularly of a more abstract nature. His thinking tends to be not so dominated or oriented by objective conditions and generally accepted ideas as the thinking of the extrovert. The thinking extrovert, however, shows a liking for overt action, and his ideas tend to be ideas of overt action. His thinking tends to be more dominated by objective conditions and generally accepted ideas than that of the introvert.

The social introvert withdraws from social contacts and responsibilities. He displays little interest in people. In contrast, the social extrovert seeks social contacts and depends upon them for his satisfaction. He is primarily interested in people.

The emotional introvert tends to repress and inhibit the outward expression of emotions and feelings. He tends not to make the typical response to simple, direct emotional appeals. On the other hand, the emotional extrovert readily expresses his emotions and feelings outwardly. He tends to make the expected response to simple, direct emotional appeals.

The second step was to formulate three distinct types of items to deal with thinking, social, and emotional reactions. The items were stated as questions to which the individual could respond by indicating whether the activity or reaction was characteristic of his or her behavior *very often, frequently, occasionally, rarely,* or *almost never.* Constructing the items was a process of translating the definitions of the three types of *I-E* into specific forms of behavior. For example, in formulating items for measuring Thinking *I-E,* an effort was made to include activities which would bring out the fundamental differences in the behavior of the introvert and extrovert as described in the definition of Thinking *I-E.* It should be emphasized that original items were devised for the inventory.

The third step in constructing the inventory was a preliminary means of making each of the three tests as homogeneous as possible. Ten experts classified the individual items in accordance with the definitions into the following six groups: Thinking Introversion, Thinking Extroversion, Social Introversion, Social Extroversion, Emotional Introversion, and Emotional Extroversion. Six of the 10 judges, i.e., a majority, agreed on the placement of 197 of the 216 items sorted. Since the 197 items were unevenly distributed among the six groups, they were supplemented by 43 new items. This first selection of items by the method of expert opinion provided a preliminary form of the inventory to be refined as described in the next paragraph.

A technique of item analysis was employed as the principal method for obtaining finally three homogeneous tests with low intercorrelation coefficients. This technique left only those items in each test which discriminated significantly with respect to the total score on that test but which were not discriminating with respect to the total score on either of the other two tests. The item analyses which determined the choice of items for the final inventory were made on the responses of two groups of college students on the preliminary forms of the tests. These two groups of students were composed respectively of 159 juniors in the College of Education at the University of Minnesota and of 160 freshmen and sophomores from three liberal arts colleges. Only the 197 of the 240 items in the preliminary inventory which had been classified by the experts were used in securing scores on each of the three tests to be used in the choice of criterion groups for the item analyses. Three in-

dices of discrimination, therefore, were computed for each item in the preliminary inventory for both groups of college students. Six item analyses were completed for all 240 items in the three tests. In other words, the discriminative values of each item in the inventory were computed for three criteria, scores on the Thinking test, Social test, and Emotional test for both the junior and freshman-sophomore samples.

The two criterion groups for each test were composed respectively of $\frac{N}{4}$ or 40 students with high scores on that test and of $\frac{N}{4}$ students with low scores on that test. The mean response to each item was determined for the two criterion groups of each test. As a measure of the discriminative value of an item, the Fisher "*t*" test of the significance of the difference in the two means was employed. If the value of "*t*" satisfied the five per cent level of significance, an item was considered to have significant discriminative power for that test.

Each item selected for the final inventory met the following standard on at least one of the two student samples: it had significant discriminating power for only one test, the test in which it had been placed by the majority of the judges. For the junior sample, 146 items met this standard in comparison with 130 items for the freshman-sophomore group. Ninety-nine items met the standard for placement in the same test with both samples. The final form of the inventory was composed of 169 items which met the standard for at least one of the two student samples.

Later, three item analyses were made from the scores of 393 College of Education seniors on the final form of the inventory, since these papers were to be used in studying the reliability and validity of the tests. One hundred and fourteen of the 169 items showed significant discriminative value in their own tests but not in the other two *I-E* measures. In the exploration of the reliability and validity of the inventory, 18 items were omitted from the final scoring. In the results of the item analyses from the senior sample, each of these items either showed more discriminating power in a test other than that in which it had been scored, or it showed significant discriminating power in two tests. The remaining 37 items which did not meet the standard were not omitted from the final scoring because they did not weaken seriously the accuracy of the

senior scores. They either were not significantly discriminating for any test or they were more discriminating in the test scored than in either of the other two tests.

There was a striking consistency in the results of the item analyses for the three samples of college students (Table 1). Of the

TABLE 1

NUMBER AND PERCENTAGE OF ITEMS OF EACH TYPE WITH SIGNIFICANT DISCRIMINATIVE VALUE FOR ONE, TWO, AND THREE SAMPLES OF COLLEGE STUDENTS

Type	Three samples No.	Three samples Per cent	Two samples No.	Two samples Per cent	One sample No.	One sample Per cent	Total
Thinking Extroversion	13	50.00	6	23.08	7	26.92	26
Thinking Introversion	27	71.05	11	28.95	0	0	38
Social Extroversion	25	80.65	2	6.45	4	12.90	31
Social Introversion	13	50.00	11	42.31	2	7.69	26
Emotional Extroversion	2	12.50	9	56.25	5	31.25	16
Emotional Introversion	2	14.29	6	42.86	6	42.86	14
Total	82	54.30	45	29.80	24	15.89	151

151 items finally scored, 84 per cent had significant discriminative value for two of the three samples, and 54 per cent for all three groups. More than 70 per cent of the thinking introvert and social extrovert items showed significant discriminating power for all three groups. No thinking introvert item discriminated significantly for fewer than two samples, and less than 10 per cent of the social introvert items met the standard for retention for only one student group. The emotional items were the weakest in the inventory, but even 69 per cent of the emotional extrovert and 57 per cent of the emotional introvert items discriminated significantly for two or more student samples.

C. INTERCORRELATION COEFFICIENTS FOR THE THREE *I-E* TESTS

This simple technique of item analysis was successful in developing tests with low intercorrelation coefficients. It will be noted from Table 2 that this technique applied to three samples of college students decreased the correlation between the Social and Emotional *I-E* tests from .54 for the original inventory to less than .25 for

TABLE 2
CHANGE IN INTERCORRELATION COEFFICIENTS FOR THE THREE *I-E* TESTS

Test	Preliminary inventory N=159	Final inventory 169 Items N=393	Final inventory 151 Items N=396	Final inventory 151 Items N=132
Thinking and Social *I-E*	+.11	—.19	—.25	—.27
Thinking and Emotional *I-E*	+.10	+.17	+.17	+.13
Social and Emotional *I-E*	+.54	+.38	+.23	+.24

the 151 items scored in the final inventory. At the same time, the intercorrelation coefficients of the Thinking test with the other two tests were raised somewhat. However, these two coefficients remained below .28 for two different groups of college students. Thus the simple technique of internal consistency coupled with absence of item discriminative power on the other criterion scores yielded three homogeneous tests with low intercorrelations. It should be noted that the intercorrelation coefficients for these tests are much lower than five of the 10 coefficients of intercorrelation reported for *An Inventory of Factors STDCR*.

With correlation coefficients as low as .27, there is no consistent tendency for individuals to be extrovert, introvert, or ambivert in all three types of *I-E* as defined and measured in this inventory. For example, an individual may have scores which indicate that he is socially extroverted, emotionally introverted, and has a medium degree of Thinking *I-E*.

The correlations shown in Table 2 between the Emotional test and the Thinking and Social tests are positive, but the correlation between the Thinking and Social tests is negative. There is a slight tendency for Thinking Introversion and Social Extroversion and for Thinking Extroversion and Social Introversion scores to be related.

D. RELIABILITY OF THE THREE *I-E* TESTS

The coefficients of reliability of the three tests were determined by both the split-half and the retest techniques. By the split-half technique, the coefficients of reliability based on the scores of 396 Education seniors were: for the Thinking test, .91; for the Social test, .88; and for the Emotional test, .75 (Spearman-Brown prophecy formula applied). The coefficients of correlation between the

test and retest scores of 101 College of Education students were: for the Thinking test, .89; for the Social test, .84; and for the Emotional test, .88. Each test had a reliability coefficient of .88 or above for one or both techniques. The tests seem sufficiently reliable for individual prediction.

E. Validity of the Three *I-E* Tests

Indirect evidence of the validity of each test was secured by determining the power of the test to differentiate known groups of college students which, on an a priori basis, would be expected to be extreme in a given type of *I-E*. The test scores of 874 University of Minnesota students were employed in this study of the differentiation of known groups.

1. *Thinking I-E Test*

The Thinking *I-E* test significantly differentiated known groups which one would anticipate to be at the extremes in that dimension. The analysis of variance technique revealed that the 395 College of Education seniors in 13 major fields were not homogeneous in Thinking *I-E*. The variance between the mean Thinking scores of the 13 major groups was significantly larger than the variance within the major groups. The mean Thinking scores of the majors in physical education, home economics, commercial education, and child welfare were extreme in the direction of extroversion; the mean scores of the majors in English, art, mathematics, social studies, and language were extreme in the direction of introversion. The differences in means of the extreme introvert and extrovert major fields were significant (Table 3). For example, the differences between the mean for the physical education majors and the means of the majors in English, in art, in social studies, and in mathematics all satisfied the one per cent level of significance. Likewise, the difference between the mean scores on the Thinking test of the commercial education majors and the art majors was highly significant.

The mean score on the Thinking test of 35 members of three engineering honorary societies was compared with the mean scores of the following four groups: (*a*) 14 mathematics majors in the College of Education, (*b*) 395 education seniors, (*c*) 30 members of two language honorary societies, and (*d*) 27 members of Phi Beta Kappa (see Table 3). In all four cases, the mean score of

TABLE 3

DIFFERENCE AND SIGNIFICANCE OF THE DIFFERENCE IN THE MEAN THINKING *I-E* SCORES BETWEEN CERTAIN GROUPS OF STUDENTS

Groups compared*	Difference in mean scores	*t*	Probability of *t*
29 Physical Education vs. 51 English Major	36.22	5.56	.01
29 Physical Education vs. 17 Art Majors	31.42	3.21	.01
29 Physical Education vs. 14 Mathematics Majors	26.62	2.96	.01
29 Physical Education vs. 50 Social Studies Majors	23.15	3.87	.01
19 Commercial Education vs. 17 Art Majors	25.64	2.89	.01
35 Engineers vs. 14 Mathematics Majors	23.05	3.64	.01
35 Engineers vs. 395 Education Seniors	12.41	2.70	.01
35 Engineers vs. 30 Members of Two Literary Honorary Societies	28.66	4.69	.01
35 Engineers vs. 27 Phi Beta Kappa Members	32.59	5.68	.01
395 Education Seniors vs. 103 Graduate Students	10.98	3.64	.01

*The first group in each case was more Thinking extroverted.

the engineers differed significantly (in the direction of extroversion) from the mean score of the other group. Likewise, the mean score of the 395 education seniors was significantly more extroverted than the mean scores of the Phi Beta Kappa members and of 103 graduate students. These differences and the differences in the major fields seem quite in accord with the definition of Thinking *I-E* in terms of which the test was constructed.

The relationship between the Thinking *I-E* test scores and the scores on a scholastic aptitude test, the Miller *Analogies Test,* was explored. A small positive correlation was found between Thinking Introversion and the Analogies scores; the computed coefficients varied from .15 to .26 with groups of 112 to 260 students in the College of Education. There was thus only a slight tendency for Thinking Introversion and high Analogies ability to be related. However, the coefficients of correlation were so low that unquestionably the Thinking *I-E* test does not measure scholastic aptitude or general intelligence.

2. *Social I-E Test*

The Social *I-E* test significantly differentiated known groups of students varying in the degree of participation in campus activities at the University of Minnesota (see Table 4). Among the senior

TABLE 4

DIFFERENCE AND SIGNIFICANCE OF THE DIFFERENCE IN MEAN SOCIAL *I-E* SCORES ON EIGHT COMPARISONS OF STUDENT GROUPS

Groups compared*	Differences in mean scores	*t*	Probability of *t*
1. 57 Sorority Women vs. 219 Non-Sorority Women	6.78	2.01	.05
2. 20 Members of Eta Sigma Upsilon vs. 207 Non-Members of Sororities or of Eta Sigma Upsilon	16.44	3.14	.01
3. 69 Eta Sigma Upsilon Members and Sorority Women vs. 207 Non-Members of Sororities or Eta Sigma Upsilon	9.06	2.89	.01
4. 9 Mortar Board Members vs. 207 Non-Members of Sororities or Eta Sigma Upsilon	17.56	2.30	.05
5. 5 Fraternity Men vs. 114 Non-Fraternity Men	10.14	.89	>.05
6. 27 Participants in Campus Activities vs. 26 Non-Participants (Women)	13.26	2.10	.05
7. 15 Participants in Campus Actvities vs. 104 Non-Participants (Men)	15.99	2.37	.05
8. 276 Women vs. 119 Men	7.39	2.79	.01

*The more extroverted group is always given first.

women in the College of Education, the mean score of the 57 members of academic sororities varied significantly in the direction of Social Extroversion from the mean score of the 219 non-members of sororities. Similarly, the mean score of the 20 members of Eta Sigma Upsilon, an organization for women leaders in the College, was significantly more socially extroverted than the mean score of the 207 senior women unaffiliated with sororities or Eta Sigma Upsilon. When 27 senior women who had at least four activities listed for them in the *Gopher Yearbook* were compared with 26 women who had either one or no activity listed, the significant difference in the mean scores indicated that the students active in

campus organizations tended to be more socially extroverted than the non-active students. A similar significant difference was obtained when the mean score of 15 senior men who were active in academic fraternities and campus organizations was compared with the mean score of the 104 remaining senior men. In summary, among the seniors in the College of Education, the members of academic sororities and fraternities, the members of Eta Sigma Upsilon, and the students active in campus organizations were significantly more extroverted than the non-affiliates and non-participants. Women students also were more socially extroverted than men students.

3. *Emotional I-E Test*

A sex difference in the scores on the Emotional test seemed logical. Social pressures and training do not encourage the young man to express outwardly such emotions as joy, sorrow, and fear as freely as the young woman. The mean scores of the women were significantly more emotionally extroverted in direction than the mean scores of the men in the following three groups: (*a*) 395 education

TABLE 5

DIFFERENCE AND SIGNIFICANCE OF THE DIFFERENCE IN THE MEAN EMOTIONAL *I-E* SCORES OF SEX AND AGE GROUPS

Groups compared*	Difference in mean scores	*t*	Probability of *t*
276 Senior Women vs. 119 Senior Men	8.75	7.29	.01
48 Freshman Women vs. 41 Freshman Men	12.34	6.08	.01
82 Summer School Women vs. 61 Summer School Men (Born before 1915)	9.33	4.71	.01
37 Freshman Women vs. 200 Senior Women	3.90	2.01	.05
37 Freshman Women vs. 23 Summer School Women (Born before 1905)	9.84	3.28	.01
200 Senior Women vs. 23 Summer School Women	5.94	2.44	.05
21 Freshman Men vs. 59 Senior Men	.60	.24	>.05
21 Freshman Men vs. 45 Summer School Men	1.18	.40	>.05
59 Senior Men vs. 45 Summer School Men	.58	.28	>.05

*The more extroverted group is always listed first.

seniors, (*b*) 89 University freshmen, and (*c*) 143 summer school students born before 1915 (see Table 5). The summer school group was composed of public school teachers doing graduate study at the University.

Age differences also seem consistent with the basic definition of Emotional *I-E*. With increasing age and greater conformity to social codes, the typical individual probably tends to become less likely to express his emotions outwardly. For women students, the sample of freshmen was significantly more extroverted than the samples of seniors and teachers born before 1905 (Table 5). The senior women were also significantly more emotionally extroverted than the group of teachers born before 1905. In each comparison, the younger group was significantly more emotionally extroverted than the older one. Similar differences in age groups were observed in samples of men students. Freshman men were more emotionally extroverted than senior men and teachers born before 1905. Likewise, the senior men were more extroverted than the teachers. However, none of the differences in male age groups was significant. Perhaps men tend to form a stable pattern of emotional reactions earlier than women.

Each test did significantly differentiate known groups of students which one would expect, a priori, to be at the extremes in a given type of *I-E*.

F. Usefulness of the *I-E* Inventory

A research study conducted in the College of Education at the University of Minnesota indicates that students' scores on the inventory should be useful to counselors and instructors. The data obtained reveal that the Thinking *I-E* test is significantly related to the scholastic achievement of students and that the Social *I-E* test is significantly associated with student teaching success. Thinking Introversion, i.e., the liking for reflective thinking, particularly of the more abstract nature, is related to high scholastic achievement. Social Extroversion, or a tendency to seek social contacts and to depend upon them for satisfaction, is characteristic of the more successful student teachers. Several research studies now planned should yield additional evidence on the usefulness of the inventory.

G. Summary

This study has reduced the confusion in the field of measurement

of *I-E* by getting away from the general undifferentiated concept of *I-E*. An inventory was constructed to measure, not a general trait, but three types or phases of *I-E* which were clearly defined. By a simple technique of item analysis, three homogeneous and relatively independent *I-E* tests were developed. Each test seems to be sufficiently reliable for individual prediction. The demonstrated ability of each test to discriminate between groups of college students which one would logically expect to be characteristically different in a given type of *I-E* justifies the conclusion that each test is sufficiently valid for the inventory to be employed in the diagnosis and counseling of college students.

References

1. Guilford, J. P., & Guilford, R. B. Personality factors, S, E, and M, and their measurement. *J. of Psychol.*, 1936, **2**, 109-127.
2. ———. Personality factors, D, R, T, and A. *J. Abn. & Soc. Psychol.*, 1939, **34**, 21-36.
3. Guilford, J. P. Manual of Directions and Norms for an Inventory of Factors STDCR. Beverly Hills, California: Sheridan Supply Company, 1940.
4. Jung, C. G. Psychological Types. New York: Harcourt, Brace, 1923. Pp. 429.

Department of Educational Psychology
University of Minnesota
Minneapolis, Minnesota