

# The Relation Between Item Format and the Structure of the Eysenck Personality Inventory

Wayne F. Velicer and John F. Stevenson  
University of Rhode Island

A Likert seven-choice response format for personality inventories allows finer distinctions by subjects than the traditional two-choice format. The Eysenck Personality Inventory was employed in the present study to test the hypothesis that use of the expanded format would result in a clearer and more accurate indication of test structure. The subjects, volunteers in a psychology course, took the standard two-choice version of the EPI and a seven-choice version one week apart, with the order counter-balanced. A principal components analysis with a varimax rotation yielded two components for the two-choice format, clearly identifiable as Eysenck's "Neuroticism" and "Extraversion" which together accounted for 18% of the variance. The seven-choice version resulted in six components accounting for 46% of the variance. The expanded format suggested inadequacies in the structure of the EPI, defined the factor structure more clearly, and explained a greater proportion of the variance. It thus demonstrated the apparent advantages of the multiple-response format for scale construction.

Traditionally, structured personality inventories have employed a two-choice item format (e.g., true-false, agree-disagree, a forced choice between two alternatives). The reasons are primarily practical: (1) ease of administration (i.e., simplicity of instruction for subjects); (2) reduced administration time; (3) ease of scoring; and (4) avoidance of scaling issues. The recent

Comrey Personality Scales (Comrey, 1970) are one of the few exceptions to this trend, employing a Likert-type seven-choice item format. The present study investigates the value of the multi-category approach, based on the reasoning that such items will permit the subject to make finer distinctions and, therefore, will provide more precise and meaningful responses. This should result in (1) increased item and scale reliability, (2) more favorable subject reactions to the inventory, and (3) a clearer and more accurate indication of the test structure.

Of these questions, the reliability issue has been the most widely researched, but conflicting results have been produced. For example, Jahoda, Deutch, and Cook (1951) and Ferguson (1941) report that reliability increases as the number of response categories increases. Bendig (1954), Komorita (1963), Peabody (1962), and Matell and Jacoby (1971) have found reliability to be generally independent of the number of response categories. Komorita and Graham (1965) found an increase in reliability with an increase in number of categories only for scales with relatively homogeneous items. Masters (1974) reported a relation only for a scale that had an initial low total score variation. It is difficult to resolve these differences because (1) different test instruments were used in the various studies and (2) different methodologies were employed.

---

*APPLIED PSYCHOLOGICAL MEASUREMENT*  
Vol. 2, No. 2 Spring 1978 pp. 293-304  
© Copyright 1978 West Publishing Co.

A potentially confounding and uncontrolled effect on reliability is the possibility that the increase in the number of response categories altered the structure of the instrument. The structure is usually determined on the two-response form of the instrument or established "theoretically." If the multiple category versions of the instruments have a different structure, then measures of internal consistency based on the scales of original (binary) form would be inappropriate, particularly if the structure is more complex. This would explain the Komorita and Graham (1965) result with respect to "heterogeneous" scales. The only study to control for structure is the recent monte carlo study by Lissitz and Green (1975), which found an increase in reliability as the number of categories increased from two to five and no change for further increases in the number of categories.

The issue of subject reaction to type of format has not been extensively investigated. Jones (1968) reported that subjects generally preferred a multi-category form to a two-choice format. This finding is supported by the first author's informal observations.

The relation between item format and the structure of the inventory has not been studied extensively. If the multi-category format does, in fact, provide meaningful finer distinctions, an analysis of structure at the item level could result in a better defined scale and inventory structure. Improvements in structure resulting from this approach would include (1) accounting for more of the total variation, (2) higher component loadings, and/or (3) identification of additional components. Joe and John (1973) employed both the traditional forced-choice (two-response) format and a six-choice response with the Rotter I-E Scale. The six-choice format resulted in two clearly interpretable factors, illustrating the potential value of the multi-category approach for probing scale structure.

In the present study, the Eysenck Personality Inventory (EPI) was selected for use in an investigation of the effects of response format on inventory structure. The EPI was originally de-

rived factorially, possesses good psychometric properties, and has been used extensively. The limited number of items (57) permits a full analysis at the item level. The reasoning in the preceding paragraph led to the expectation that the standard EPI with two-choice format should yield two components (or three if the Lie Scale emerges as a separate component), while a Likert format version of the EPI should yield additional components and/or better defined components; these components should then account for more of the total variance.

### Method

The subjects were students enrolled in two sections of a lower level psychology course. Participation was voluntary, and the students were told the general nature of the task, but not the specific hypothesis. Each student completed the standard Eysenck Personality Inventory and a seven-choice version of the EPI. The seven-choice version was developed by employing the item format of the Comrey Personality Scales with the 57 items of the EPI. Comrey employed two different scalings for different items: one based on a frequency concept (ranging from "Always" to "Never") and the other based on a likelihood concept (ranging from "Definitely" to "Definitely Not"). This method was easily adapted to the items of the EPI. Administration was in two sessions one week apart. Approximately half the students took the Standard EPI first and then the Likert EPI, while the order was reversed for the remaining students. Only students who completed both forms were included in the final sample ( $N = 77$ ).

For each version of the inventory, a principal components analysis was performed on the  $57 \times 57$  matrix of item intercorrelations. Velicer's (1976) Minimum Average Partial (MAP) correlation method was used to determine the number of components to extract. A varimax rotation was performed on the component pattern.

A critical issue in this study is the method of

Table 1  
 Comparison of Results of Component Analysis with EPI Scoring Keys.\*

Item	Varimax Rotated Component Structure		Theoretical Structure		
	Components		Scales		
	1	2	Neuroticism	Extraversion	Lie
1	.43	.05		x	
2	.49	-.07	x		
3	-.17	.34		x	
4	.23	.09	x		
5	-.19	-.13		x	
6	.15	.09			x
7	.58	-.18	x		
8	.31	.07		x	
9	.56	-.07	x		
10	.22	.12		x	
11	.36	-.29	x		
12	.34	.18			x
13	.18	.26		x	
14	.53	-.18	x		
15	.07	-.44		x	
16	.61	-.20	x		
17	-.11	.65		x	
18	.04	-.02			x
19	.44	.02	x		
20	.23	-.12		x	
21	.47	.08	x		
22	.40	.21		x	
23	.57	.10	x		
24	-.16	-.18			x
25	.01	.32		x	
26	.50	-.14	x		
27	-.07	.59		x	
28	.41	-.07	x		
29	.04	-.72		x	
30	.11	.17			x
31	.45	-.13	x		
32	.07	-.33		x	
33	.51	.10	x		
34	.25	-.02		x	
35	.40	.25	x		

(Continued on next page)

Table 1: Continued

Varimax Rotated Component Structure			Theoretical Structure		
Item	Components		Scales		
	1	2	Neuroticism	Extraversion	Lie
36	.33	-.11			x
37	-.10	-.45		x	
38	.29	-.27	x		
39	-.17	.08		x	
40	<u>.62</u>	-.12	x		
41	-.08	-.02		x	
42	-.01	.02			x
43	.19	.00	x		
44	.21	<u>.32</u>		x	
45	<u>.30</u>	.10	x		
46	<u>.33</u>	.18		x	
47	<u>.44</u>	-.15	x		
48	.26	.16			x
49	-.29	<u>.52</u>		x	
50	<u>.49</u>	-.11	x		
51	.08	-. <u>68</u>		x	
52	<u>.36</u>	-. <u>36</u>	x		
53	-.19	<u>.43</u>		x	
54	.05	.24			x
55	.27	-.04	x		
56	.15	<u>.37</u>		x	
57	<u>.39</u>	-.20	x		
Eigen Value	6.6446	3.9376			
Percent of Variance	11.66	6.91			

\*Loadings greater than .30 are underlined.

determining the number of components to be extracted. The MAP method (Velicer, 1976) is rather new, but it possesses a number of advantages for a study of this type. The procedure determines the number of components by successively partialing out components until the average squared partial correlation reaches a minimum. Components retained by this procedure are clearly "common" components. For com-

parison purposes, an exact stopping point that is empirically determined is necessary. Alternative, more traditional methods, such as the scree test or the eigenvalue-greater-than-one criterion, give less satisfactory results. For both analyses, the scree test provided no clear solution. The eigenvalue-greater-than-one method resulted in 21 components for the binary format and 17 for the Likert format. However, in both cases, the

Table 2

## Cross Classification Table for Scale and Components

Scale	Component		Unclassified	Totals
	1	2		
Neuroticism	19	0	5	24
Extraversion	4	13	7	24
Lie	2	0	7	9
Totals	25	13	19	57

additional components were either poorly defined components or "unique" components.

### Results

The analysis of the traditional two-choice version of the Eysenck Personality Inventory resulted in two components by the MAP method (see Table 1). These were clearly identifiable as "Neuroticism" and "Extraversion." However, the two components together accounted for only 18% of the total variance. An analysis of the varimax-rotated pattern showed that 19 items did not load on either component and were unclassifiable (see Table 2).

The seven-choice version of the EPI resulted in six components (see Table 3). The six components together accounted for 46% of the total variance, with the first two components alone accounting for 26% of the variance. If six components had been extracted for the two-choice version, a total of 37% of the variance would have been accounted for. The first two components were identified as "General Anxiety" and "Social Extraversion." The four new components were identified as "Compulsivity," "Impulse Control," "Health Concerns," and "Affiliative Concern." Each component was identified by at least four different items with loadings of .49 or higher. Table 4 lists marker items for each of these six components.

The pool of items represented in Table 4 was employed to calculate a scale score for each of the six components from the Likert format.

Items with negative loadings were reflected, and the unweighted sum of the items formed the scale score. The binary version was scored in the standard manner. The means, standard deviations, and intercorrelations for the three traditional scales and six new scales are presented in Table 5.

An examination of the correlations suggests a match of the Extraversion and Social Extraversion, Neuroticism and General Anxiety, and Lie and Compulsivity Scales. Since Health Concerns and General Anxiety correlated with Neuroticism and with each other, there is some support for Neuroticism as a second order factor. Impulse Control and Affiliative Concerns were essentially uncorrelated with the other seven scales.

### Discussion

These results have implications for several different issues. The appropriateness of the present form of the EPI is a relatively specific issue raised by the data and will be discussed first.

The extensive examination of the factor structure of the EPI by Eysenck and Eysenck (1969) fails to report the percent of total variance accounted for by factors in any of the several analyses cited; such figures are also absent from a more recent investigation of the EPI (Howarth, 1976). Hence the 18% figure obtained in the present study must stand as the only available estimate of the variance accounted for by Eysenck's two factors when the two-choice response format

Table 3  
 Varimax Rotated Component Pattern  
 for Likert Item Version of EPI\*

Item	Component					
	1	2	3	4	5	6
1	.48	.17	-.01	-.11	.10	-.32
2	.65	.00	.07	.26	.09	.10
3	-.36	.44	-.28	.22	-.24	-.17
4	-.21	.05	.04	-.50	.17	-.09
5	-.20	.06	.47	.36	-.08	-.24
6	.06	.12	.52	-.10	-.17	-.14
7	.48	-.31	.09	.43	.28	-.09
8	.15	-.03	-.24	.68	.07	-.13
9	.48	-.16	-.25	.39	.37	-.04
10	.00	.09	-.32	.42	.22	-.01
11	.44	-.31	.12	.22	.23	-.04
12	-.18	.11	-.14	-.72	-.01	-.26
13	-.08	.48	-.32	.41	.26	-.11
14	.65	.04	.12	.39	.15	-.14
15	-.05	-.58	-.05	-.09	-.01	.19
16	.62	-.03	.15	.30	.04	.32
17	.08	.51	-.16	-.08	.00	-.43
18	.42	.02	.17	.10	-.19	-.15
19	.36	.05	.19	.31	.35	-.18
20	.08	.02	-.21	.30	-.21	.08
21	.66	.14	-.04	.13	.33	-.08
22	.00	.06	.03	.68	.04	-.41
23	.77	.00	.12	.21	-.07	-.12
24	-.38	-.13	.50	.06	.08	.31
25	-.18	.52	.10	.03	.16	.28
26	.19	-.24	.53	.16	.37	-.13
27	-.07	.76	.17	.16	-.17	-.07
28	.22	-.02	.14	-.04	.55	-.10
29	-.23	-.55	-.33	.13	-.27	.35
30	-.22	-.02	.11	-.29	.04	-.52
31	.28	-.29	.03	.04	.65	.06
32	-.03	.41	.46	-.04	.30	.10
33	.08	.18	.17	.21	.56	-.16
34	-.09	.03	-.37	-.02	-.09	.24
35	.45	-.11	-.04	-.16	.40	.20
36	.02	-.08	.21	.00	.22	.42
37	-.14	.12	.05	-.03	.07	.59

(Continued on next page)

Table 3: Continued

Item	Component					
	1	2	3	4	5	6
38	.23	-.24	-.16	<u>.61</u>	.22	.03
39	<u>.40</u>	.11	.13	<u>.10</u>	.10	-.13
40	<u>.54</u>	-.11	.26	<u>.31</u>	<u>.37</u>	.06
41	-.01	.07	-.55	.08	-.04	.17
42	-.06	.07	-.51	-.10	.09	-.03
43	<u>.56</u>	.01	<u>.12</u>	-.13	.14	-.21
44	<u>.08</u>	<u>.37</u>	.29	-.02	<u>.35</u>	-.18
45	.10	<u>.15</u>	-.06	-.02	<u>.62</u>	.12
46	.08	.22	-.06	.07	<u>.17</u>	<u>.51</u>
47	<u>.44</u>	-.15	<u>.31</u>	.19	<u>.38</u>	<u>.01</u>
48	-.02	-.08	<u>.03</u>	-.45	<u>.13</u>	-.11
49	-.63	<u>.31</u>	.14	<u>.00</u>	.04	-.07
50	<u>.51</u>	<u>.03</u>	.26	<u>.41</u>	.07	-.22
51	-.29	-.62	-.11	-.20	.12	<u>.39</u>
52	<u>.70</u>	-.33	-.01	.02	-.05	<u>.01</u>
53	-.14	<u>.77</u>	.07	-.11	.07	.06
54	-.26	<u>.32</u>	.22	.01	-.20	-.08
55	-.03	<u>.03</u>	.12	.05	<u>.61</u>	.00
56	.00	.03	.07	.21	-.19	-.49
57	<u>.55</u>	-.31	.13	.03	.23	<u>.00</u>

Eigen  
Value 10.1486 4.8380 3.5885 3.2378 2.6728 2.1825

Percent  
of  
Variance .1780 .0849 .0630 .0568 .0490 .0383

\*Loadings greater than .30 are underlined.

is used. This low figure and the large residue of unclassifiable items calls for further investigation and tentatively points to the need for revision of the inventory. The meaningfulness of neuroticism and extraversion scores derived according to the test manual is also called into question.

Use of the expanded response format has the apparent advantage of producing components

which account for a much higher proportion of total variance. Even when only the first two components are considered (those which most clearly represent Eysenck's Neuroticism and Extraversion factors), 26% of the total variance was accounted for. These results indicate that users of the EPI might obtain more meaningful scores on Neuroticism and Extraversion by basing them on the items which load highest on these two components.

Table 4  
 Marker Variables for Varimax Rotated Components of the Seven  
 Choice Item Format Version of the EPI.

Loading	Test Item Number	Item	EPI Key
Component 1. "General Anxiety"			
.77	23	Are you often troubled about feelings of guilt?	N
.70	52	Are you troubled with feelings of inferiority?	N
.66	21	Do you daydream a lot?	N
.65	14	Do you often worry about things you should not have done or said?	N
.65	2	Do you often need understanding friends to cheer you up?	N
-.63	49	Would you say you were fairly self-confident?	E
.62	16	Are your feelings rather easily hurt?	N
.56	43	Do you have many nightmares?	N
.55	57	Do you suffer from sleeplessness?	N
.54	40	Do you worry about awful things that might happen?	N
.51	50	Are you easily hurt when people find fault with you or your work?	N
Component 2. "Social Extraversion"			
.77	53	Can you easily get some life into a rather dull party?	E
.76	27	Do other people think of you as being very lively?	E
-.62	51	Do you find it hard to really enjoy yourself at a lively party?	E
-.58	15	Generally do you prefer reading to meeting people?	E
-.55	29	Are you mostly quiet when you are with other people?	E
.52	25	Can you usually let yourself go and enjoy yourself a lot at a gay party?	E
.51	17	Do you like going out a lot?	E

(Continued on next page)

Table 4: Continued

Loading	Test Item Number	Item	EPI Key
Component 3. "Compulsivity"			
-.55	41	Are you slow and unhurried in the way you move?	E
.53	26	Would you call yourself tense or "high-strung?"	N
.52	6	If you say you will do something do you always keep your promise, no matter how inconvenient it might be to do so?	L
-.51	42	Have you ever been late for an appointment or work?	L
.50	24	Are all your habits good and desirable ones?	L
Component 4. "Impulse Control"			
.72	12	Once in a while do you lose your temper and get angry?	L
.68	8	Do you generally do and say things quickly without stopping to think?	E
.68	22	When people shout at you, do you shout back ?	E
.61	38	Are you an irritable person?	N
Component 5. "Health Concerns"			
.65	31	Do ideas run through your head so that you cannot sleep?	N
.62	45	Are you troubled by aches and pains?	N
.61	55	Do you worry about your health?	N
.56	33	Do you get palpitations or thumping in your heart?	N
.55	28	After you have done something important, do you often come away feeling you could have done better?	N
Component 6. "Affiliative Concern"			
.59	37	Do you hate being with a crowd who plays jokes on one another?	E
-.52	30	Do you sometimes gossip?	L
-.51	46	Would you be very unhappy if you could not see lots of people most of the time?	E
-.49	56	Do you like playing pranks on others?	E

Table 5  
Means, Standard Deviations, and Intercorrelations for the Scale Scores

Scales	Mean	Standard Deviation	1	2	3	4	5	6	7	8
1 Extraversion	12.42	3.46								
2 Neuroticism	9.65	5.09	-.10							
3 Lie	2.78	1.62	-.28	.02						
4 General Anxiety	40.53	8.12	-.06	.76	.04					
5 Social Extraversion	33.62	5.96	.67	-.21	-.14	-.26				
6 Compulsivity	22.30	4.13	-.10	.19	.52	.18	-.04			
7 Impulse Control	14.71	2.35	.19	.28	-.24	.26	-.03	-.15		
8 Health Concerns	17.14	4.26	-.03	.55	.12	.42	-.06	.14	.18	
9 Affiliative Concerns	14.17	3.45	.35	-.09	-.26	-.11	.27	.03	-.05	-.03

The four additional components which emerged with the multi-choice format may have some value in their own right. However, extensive research would be needed before these components could be viewed as established. Additional items would be needed, and the relation to previously established personality scales of other inventories would have to be investigated. The component names should be viewed as tentative working names only.

Limited support for the component structure which emerged in the present study may be found in Guilford's (1975) critical analysis of the Eysenck Personality Inventory. He suggested that the Extraversion-Introversion Scale does not represent a factor at any level, but is rather a "shotgun wedding" of two first-order factors, impulsivity and sociability. The present study did extract a Social Extraversion component and an Impulse Control component, consisting primarily of items drawn from the Extraversion Scale. However, there was also a third component, Affiliative Concern, comprised of *E*-scale items. The present study found the Neuroticism scale items also split into two (correlated) components; this somewhat parallels the substructure anticipated by Guilford, who viewed Neuroticism as a legitimate second-order factor.

A more general question toward which the present study is directed concerns the value of a finer-grained response format in the development of personality scales and inventories. The results indicate that a more precise definition of scales and a greater explained portion of the total variance follow from use of the expanded format. Research employing scales constructed in this way would be more likely to obtain meaningful relationships to other variables, and the items involved in these relationships would be more accurately identified. The results also support the hypothesis that conflict in the research on the relation between item format and reliability is due to changes in structure. Since reliability as measured by coefficient alpha is directly related to the size of the eigenvalues, the reliabilities of the components from the seven-choice format is clearly better.

Two additional methodological issues must be considered. The obtained results might have occurred if either the estimates of the correlation coefficients for the binary format were underestimated or the estimates for the Likert format were inflated. In the first case, the substitution of tetrachoric  $r$ 's for the phi coefficients employed would result in higher values in the correlation matrix. This also would result in 22 negative eigenvalues. It seems more reasonable to view the (potentially) lower correlation values resulting from the use of the phi coefficient as accurately reflecting a problem implicit in the use of binary data; therefore a purely statistical correction need not be attempted. In the second case, inflated values in the Likert format data could occur as a result of an extremity response set. Such a response set would result in increased item standard deviations. The average standard deviation for the 57 items was 1.29. While this was slightly larger than might be expected if each item were normally distributed, it can be accounted for by the presence of a number of skewed items. Likewise, a review of the original responses by subject does not support the presence of an extremity response set.

A number of limitations restrict the strength of generalizations from this study. First, sample size is somewhat small. Second, a critical decision, regarding the number of components which should be extracted, is based on a relatively new criterion. Third, generalizations about the advantages of the multiple-response format require replications using other personality inventories; of particular interest would be the effect of a format change on inventories which have been developed by different procedures, such as theoretical approaches or first-order factor analytic procedures.

With the above limitations in mind, the following conclusions may be drawn from this study. The present scoring procedure of the binary form of the Eysenck Personality Inventory is acceptable, but inefficient. The factor structure is more clearly defined with the multiple-response format and more variance is explained.

The Likert format generally supports Guilford's (1975) position. More generally, the multiple-response format has demonstrated very good potential for improving the quality of personality inventories. Potential benefits include greater scale reliability, a more clearly defined factor structure, and more favorable subject reactions.

## References

- Bendig, A. W. Reliability and the number of rating scale categories. *Journal of Applied Psychology*, 1954, 38, 38-40.
- Comrey, A. L. *EITS Manual for the Comrey Personality Scales*. San Diego: Educational and Industrial Testing Service, 1970.
- Eysenck, H. J., & Eysenck, S. B. G. *Personality structure and measurement*. San Diego: Knapp, 1969.
- Ferguson, L. W. A study of the Likert technique of attitude scale construction. *Journal of Social Psychology*, 1941, 13, 51-57.
- Guilford, J. P. Factors and factors of personality. *Psychological Bulletin*, 1975, 82, 802-814.
- Howarth, E. A psychometric investigation of Eysenck's personality inventory. *Journal of Personality Assessment*, 1976, 40, 173-185.
- Jahoda, M., Deutsch, M., & Cook, S. W. (Eds.). *Research methods in social relations*. New York: Dryden Press, 1951.
- Joe, V. C., & John, J. C. Factor structure of the Rotter I-E Scale. *Journal of Clinical Psychology*, 1973, 29, 66-68.
- Jones, R. R. Differences in response consistency and subjects' preferences for three personality inventory response formats. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, 3, 247-248.
- Komorita, S. S. Attitude content, intensity, and the neutral point on a Likert scale. *Journal of Social Psychology*, 1963, 61, 327-334.
- Komorita, S. S., & Graham, W. K. Number of scale points and the reliability of scales. *Educational and Psychological Measurement*, 1965, 4, 987-995.
- Lissitz, R. W., & Green, S. B. Effect of the number of scale points on reliability: A monte carlo approach. *Journal of Applied Psychology*, 1975, 60, 10-13.
- Masters, J. R. The relationship between number of response categories and reliability of Likert-type questionnaires. *Journal of Educational Measurement*, 1974, 11, 49-53.

- Matell, M. S., & Jacoby, J. Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement*, 1971, *31*, 657-674.
- Peabody, D. Two components in bipolar scales: Direction and extremeness. *Psychological Review*, 1962, *69*, 65-73.
- Velicer, W. Determining the number of components from the matrix of partial correlations. *Psychometrika*, 1976, *41*, 321-327.

### Acknowledgments

*An early version of this paper was presented at the Spring 1977, Eastern Psychological Association Meeting, Boston. The authors acknowledge, with deep thanks, the assistance of Raymond Kilduff in the initial stages of the study.*

### Author's Address

Wayne F. Velicer, Department of Psychology, University of Rhode Island, Kingston, RI 02881.