

A Psychometric Investigation of the Eysenck Personality Questionnaire

Edward Helmes

The University of Western Ontario and London Psychiatric Hospital

Summaries of item properties for the Eysenck Personality Questionnaire (EPQ) are reported, based upon the responses of 99 male and 92 female undergraduate psychology students. Means, standard deviations, and internal consistencies were comparable to those already published, with the exception of lower reliabilities for the P and L scales. A high incidence of extreme p values and correlation with social desirability was noted for the P scale. Item factor analyses at both the first- and third-order levels did not replicate the four-factor structure claimed by Eysenck; at best, four components accounted for approximately 30% of the observed variance. Many items did not load in accordance with the EPQ scoring key, and many items did not load on an interpretable component. The question of difficulty factors is discussed: None were located. It is concluded that there is little empirical support for the structure claimed for the EPQ.

The Eysenck Personality Questionnaire (EPQ) is a major revision of previous Eysenck scales. The modifications consist of the addition of a 25-item scale to assess Psychoticism (P), or "toughmindedness," and alterations to the older Neuroticism, or emotionality (N; 23 items), Extraversion (E; 21 items), and Lie (L; 21 items) scales. Eysenck and Eysenck (1975) stated that "the E and N scales of the present questionnaire

are so similar to the corresponding scales of the other questionnaires that whatever has been discovered about correlates of E and N with the use of the older scales must be assumed to apply with equal force to the new scales" (p. 3).

Such invariance of scales over a revision of the test is reassuring to those many users of the Eysenck Personality Inventory (EPI; Eysenck & Eysenck, 1968) or the Maudsley Personality Inventory (MPI; Eysenck, 1962). However, a close examination of the items of the EPQ and EPI provides a somewhat different picture. In fact, only 25 of the 57 EPI items appear on the EPQ (8 E, 4 L, and 13 N); and 8 of those 25 have slightly altered wording. The N scale is the most similar to the old form, but the addition of so many new items calls into question the claimed similarity between old and new. It is possible, however, that a lower resemblance of the EPQ to the EPI is advantageous. A recent investigation by Howarth (1976) found the claims for both E and N scales to be somewhat exaggerated, with the E scale being more poorly structured than the N scale.

Howarth (1976) has reviewed the origins of the EPI in detail and has pointed out that item factor analyses of the EPI were not reported until after the scales had been formed (Eysenck & Eysenck, 1969). The EPQ manual (Eysenck & Eysenck, 1975) states that the P, E, and N scales were formed as a result of "a lengthy series of

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 4, No. 1 Winter 1980 pp. 43-55
© Copyright 1980 West Publishing Co.

about 20 factorial studies . . ." (p. 5). Changes were made in order to reduce correlations among factors, to eliminate items loading on more than one factor, and to increase scale reliability. No factor analytic results are reported in the manual; such results are reported in a later publication (Eysenck & Eysenck, 1976).

It is the purpose of this paper to report on some basic item properties of the EPQ for Canadian university students and to evaluate critically the structural properties of the EPQ scales. This latter effort may be seen as both an extension and replication of Howarth's (1976) study.

Method

Subjects

Respondents were 191 introductory psychology students (92 males and 99 females) at The University of Western Ontario, who received course credit for the required research participation. During three weekly 1-hour testing periods, respondents also completed several other personality tests. The EPQ was completed in the final period. No respondent was included in the sample if more than two items were omitted. Any omitted item was scored as a false response.

Procedure

Most analyses were conducted on the entire sample. Howarth (1976) reports results separately for males and females. With the smaller sample in the present case, it was thought wiser to deal with a larger group than to explore the possibility of sex differences with samples of marginal size. In addition, inspection of Howarth's (1976) results show that in most cases, the differences between the sexes are not major. Although scale means are most discrepant, means are not used in correlational analyses and can be disregarded in most cases. It remains possible, however, that sex differences in scale means are reflected in the factor structure. Eysenck and Eysenck (1976, p. 53) stated that "similar if not identical factors" can be derived from analyses of groups distinguished by sex. This supports the present analyses.

Results

Item Analyses

Table 1 reports means, standard deviations, and KR-20's for the four EPQ scales. Means and standard deviations are within approxi-

Table 1
Statistics for Scales

Scale		Number of Items	Mean	Standard Deviation	KR-20	A	B
Psychoticism:	M		4.0	2.7	.59		
	F		2.4	1.9	.45		
	Total	25	3.2	2.4	.59	.08	.02
Extraversion:	M		14.8	4.0	.79		
	F		14.1	4.9	.87		
	Total	21	14.5	4.5	.84	.32	-.05
Neuroticism:	M		10.4	4.8	.83		
	F		12.7	5.2	.87		
	Total	23	11.6	5.1	.86	.49	-.07
Lie:	M		6.4	3.6	.75		
	F		6.8	3.6	.74		
	Total	21	6.6	3.6	.75	-.22	.06

Note: M indicates males, F, females.

A = Mean item-total correlation for keyed scale.

B = Mean item-total correlation for irrelevant scale.

Table 2
Frequency of extreme *p*-values

Range	Scale			
	P	E	N	L
<.2 or >.8	22 (88%)	6 (29%)	5 (22%)	9 (43%)
<.1 or >.9	13 (52%)	1 (5%)	0 (0%)	1 (5%)

mately .5 of the normative values for the 20-29 year age group in the manual and of those reported by Loo (1979). Internal consistencies of the E and N scales are close to values of the normative group, but both L and P scales showed lower internal consistency. The drop for the P scale was particularly sharp, dropping to .45 for females. This may be attributable to the fact that the P scale is the newest and least developed EPQ scale. (The substantially lower reliability of P has also been noted by Loo, 1979.)

Table 2 reports the number of items with extreme *p* values for the four EPQ scales, with percentage of items keyed on the scale in parentheses. Howarth (1976) found only 21% of E, 17% of N, and 22% of L items were endorsed by fewer than 20% or more than 80% of the respondents to the EPI. The figures for the EPQ are roughly equivalent for the E and N scales, but substantially higher for the L scale, and extremely high for the P scale. On the P scale *p* values ranged from .02 (Item 26) to .97 (Item 61).

Items with extreme *p* values are undesirable for several reasons. First, they discriminate poorly except for individuals at the extremes, who are, in fact, a minority of respondents.

Second, they contribute relatively little to scale internal consistency. Third, they have maximum correlations with other items or variables that are less than unity, causing difficulty in interpretation and biasing analyses based on correlations, such as factor analysis. Fourth, items with extreme *p* values may attract the use of stylistic or deviant responses.

Social Desirability

Social desirability has long been identified as a major determinant of item responses to personality items (Berg, 1967; Edwards, 1970; Rogers, 1971). An earlier study by Farley (1966) reported a negative correlation of N scores on the EPI with an Edwards social desirability scale; the E and L scales did not correlate. In order to evaluate the role of social desirability in the EPQ, the scale scores of the EPQ were correlated with the Desirability scale of the Personality Research Form, Form E (PRF-E; Jackson, 1974). In addition, the Differential Reliability Index (DRI; Jackson, 1970) was calculated for each item in order to evaluate the role of desirability at the item level. Table 3 presents the mean DRI for

Table 3
Results of Desirability Analysis

	Scale			
	P	E	N	L
No. Omitted Items	3	0	0	0
Mean DRI	.27	.46	.47	.38
Correlation with PRF-E Desirability	-.36	.24	-.58	.32

Note: All values of the correlations with desirability are significantly different from zero at the .01 level.

the EPQ scales, as well as the correlations of the scales with the PRF-E Desirability scale. Three items of the P scale correlated more highly with desirability than with P and were omitted from this analysis.

The DRI and correlations of scale scores with desirability provide distinctly different bits of information about the role of desirability in a test. Judgments of item desirability provide a third piece of information. The DRI reflects the relative importance of scale content and desirability content in an item. On these grounds, the P scale was the most affected by desirability, Items 18, 22, and 37 reflecting more desirability content than Psychoticism content. The other scales were affected to a lesser degree, although the influence of desirability was fairly strong. A correlation of a scale score with desirability is more indicative of the relative desirability of the trait itself. Thus, both the P and the N scales are undesirable, and the other scales are desirable. Eysenck and Eysenck (1975) interpreted a high score on the L scale as indicating a tendency to "fake good." The positive correlation with the PRF-E Desirability scale is therefore understandable.

Item Factor Analyses

To date, there have been few reports of factor analyses of the EPQ. Loo (1979) has reported one such analysis, and Eysenck and Eysenck (1976) have reported on the factor analyses conducted during the construction of the EPQ.

Two slightly different methods were used to evaluate the degree to which the EPQ items define the factors. The first involved a components analysis of the 90 EPQ items plus the PRF-E Desirability scale, followed by Varimax (Kaiser, 1958) and Promax (Hendrikson & White, 1964) rotations. A Procrustes rotation (Schöneman, 1966) to the EPQ scoring key was also carried out. The second method was essentially the same as that used by Howarth (1976) as the method approved by Eysenck. This involved a hierarchical procedure consisting of an initial compo-

nents analysis, retaining all components with eigenvalues exceeding unity.

These components were rotated using Varimax and Promax. The correlation matrix of rotated components from the Promax rotation was then decomposed, and second-order components with eigenvalues exceeding unity were retained and rotated as for the first-order components. Loadings of items upon the second-order components were calculated for the Varimax and Promax rotations of second-order components. The correlations of second-order components from the Promax rotation were decomposed, and the process was repeated, retaining four components at the third-order level. Item loadings were then calculated for the third-order components for both Varimax and Promax rotations.

The item correlation matrix had 2,439 unique positive entries and 1,656 negative entries. The mean interitem correlation was .028, with a standard deviation of .122. The mean absolute correlation was .095, with a standard deviation of .081. The lowest correlation was $-.549$; and the highest, .644.

The first method revealed that four components accounted for only 23.2% of the variance of the original correlation matrix. The majority of the loadings in all three rotated factor pattern matrices were quite low, generally below .3. The matrices were interpreted in terms of the number of items that had their highest loadings on the scale on which they were keyed. (This information is given in Table 4.)

In general, all three rotations gave roughly similar results. No scale had all its keyed items load highest. In this regard, the N scale was the best, with 22 of 23 keyed items loading highest, and the P scale was the poorest, with 14 items of 25 having the highest loadings on their keyed scale. The E and L scales were intermediate. All scales showed relatively little change from Varimax to Promax rotations. This tends to support Eysenck's claim of orthogonality of these dimensions. Finally, the failure of the Procrustes rotation to increase substantially the number of

Table 4
Number and Percentage of Items with Highest
Loading on their Keyed Scale

Scale	Number of Items	Rotation		
		Varimax	Promax	Procrustes
P	25	14 (56%)	14 (56%)	14 (56%)
E	21	17 (81%)	16 (76%)	18 (86%)
N	23	22 (96%)	22 (96%)	22 (96%)
L	21	18 (86%)	18 (86%)	17 (81%)

items with their highest loading on their keyed scale would argue that the item factor structure of the EPQ is not entirely consistent with its keying. It is normally to be expected that an item on a scale will have its highest loading on that scale, as has been demonstrated for the PRF (Helmes & Jackson, 1977).

Table 5 gives the mean loadings of items keyed on scales for the Procrustes rotation. One E item (Item 2) failed to load in the keyed direction. All other items did so. As would be expected, keyed items loaded more highly on the targeted component than did nonkeyed items. However, the mean loading for the P scale was less than .3, implying a rather weak correspondence to the keyed structure. The other three scales were more satisfactory. The means for all scales of nonkeyed items, however, averaged .11, which is rather high if the items do indeed reflect only four univocal components. This matter will be explored in more detail by means

of analyses of more than the first four components.

In fairness to Eysenck, it must be emphasized that P, E, and N are claimed to exist as "superfactors" (Eysenck, 1978) and that their appearance is best seen at the second- or third-order level. At the same time, this invariably also means that an even smaller proportion of the total variance will be accounted for by the higher order factors, diminishing their importance as useful constructs for explaining behavior.

First-Order Analysis

For the second analysis, 31 components had eigenvalues in excess of 1, accounting for 70.3% of the variance. These eigenvalues were as follows: 8.40, 5.97, 3.54, 3.21, 2.55, 2.51, 2.25, 2.15, 2.05, 1.96, 1.90, 1.86, 1.78, 1.70, 1.67, 1.59, 1.53, 1.51, 1.44, 1.42, 1.35, 1.30, 1.29, 1.22, 1.21, 1.20, 1.16, 1.12, 1.09, 1.04, and 1.01. Results differed

Table 5
Mean Absolute Factor Loading of Scale and Non-scale
Items in Procrustes Rotated Matrix

Scale	Mean Loading of Scale Items	Mean Loading of Non-Scale Items
Psychoticism	.255	.123
Extraversion	.455	.095
Neuroticism	.472	.115
Lie	.377	.095

Note: False-keyed items have been reflected.

slightly across the two different rotations but were broadly similar. Only items loading in excess of $|.4|$ were interpreted. In the Varimax rotation, 13 components had only one salient item, with a maximum of 10 items loading on the second component. Components in the Promax rotation had from 1 to 9 salient items. There were 20 items that were not salient on any component in the Varimax solution, and 7 that were not salient in the Promax rotation. Because of the general similarity of the results of the two rotations, they will be discussed together. Components with only 1 or 2 salient items were not interpreted. The following section discusses only those 9 components with 3 or more salient items; these components accounted for 31.1% of the variance. The 22 components not analyzed accounted for 39.2% of the variance. The numbers refer to the order of extraction of the components, and negative loadings are indicated by a minus sign preceding the item number. Unless indicated, items were salient in both rotations.

I. Salients -15, -19, -72, and -80. Also -27 and -34 in Varimax only. 10.8% of total variance. This component appears similar to the negative pole of the PRF Defence scale, with items relating to a lack of irritability and to sensitive feelings. This is, in some ways, similar to the worrying aspect of Eysenck's "neuroticism." All items are keyed on the N scale. This component is similar to Loo's (1979) Factor III, which he termed Social Sensitivity.

II. Salients -5, -10, -14, 21, -40, 42, -45, -70, and -86. Also 82 on Varimax only. 7.1% of total variance. All items are related to social activities and partying. This is clearly a sociability component, related to Extraversion. All items are keyed on the E scale. This component shares 7 items with Loo's (1979) Factor I, Sociability.

IV. Salients 4, 16, 59, and 63. 3.8% of total variance. All items are keyed on the L scale and involve some aspect of mildly dishonest behavior, such as cheating or taking advantage of someone. This is Loo's (1979) Factor VI.

VI. Salients -17 and -52. Also 56 on Promax only. 3.0% of total variance. All three items are from the E scale. This component appears to

be another sociability component. Item content deals with meeting people or mingling socially.

VII. Salients -31, -41, and -75. 2.7% of total variance. All items are keyed on the P scale and involve nervous tension or health concern. All three items load on Loo's (1979) Factor V, Nervous Tension.

VIII. Salients 8, 44, and -76. 2.6% of total variance. Two items are from the L scale and one is from the P scale. Although this component is difficult to characterize, it appears to be based on self-aggrandizement.

XXII. Salients 3, 7, 12, 23, 66, and -Desirability. 1.6% of total variance. This component includes items on moodiness, depression, and anxiety, all associated with mild neuroticism. All items are keyed on the N scale. This component shares four items with Factor II, Mood Fluctuations, reported by Loo (1979).

XXIII. Salients -20, 88, 89. 1.5% of total variance. Two items are from the L and one from the N scale. This component is similar to the first, without the aspect of worry. It involves touchiness and unwillingness to admit mistakes.

XXIV. Salients -25, 29, -49. 1.5% of total variance. All items are again keyed on the E scale. This component also seems to be associated with social activity.

Seven of these nine components all clearly appear associated with aspects of Eysenck's scales; the other two are mixed. However, it must be kept in mind that these results involve at most 41 of the 90 EPQ items. Together, the 9 components accounted for only 31.1% of the variance. In addition, only 5 components replicated factors reported by Loo (1979). The differences between these results and those of Loo may, in part, be attributable to the smaller sample in this case, with a resulting lower degree of stability. However, the major factors should be strong and should appear despite the relatively small number of subjects.

Second-Order Analysis

The correlation matrix of the 31 Promax-rotated components (Table 6) was subjected to a

components analysis. Eight components had eigenvalues in excess of 1, accounting for 67.4% of the variance of the reduced correlation matrix (47.4% of the total variance). The eigenvalues were 5.23, 3.80, 2.89, 2.49, 2.13, 1.69, 1.47, and 1.18. The loadings of the items on the second-order (S) components were calculated and interpreted as before. Fifty items did not load on any component in the Promax rotation using the .4 criterion, while 24 did not load in the Varimax rotation. Thirty-six items had loadings on two or more components in the two respective rotations. The interpretation was based upon loadings of items on the second-order components because of the small proportion of interpretable first-order components. Unless otherwise indicated, items were salient in both rotations. Reliable variance is interpreted as the 70.3% of total variance accounted for by the 31 largest components.

S1. Salient 24. Items 10, 32, -34, 36, -38, -47, 52, -66, -72, -73, 82, and 86 only on the Varimax rotation. 16.9% of second-order variance and 11.9% of reliable variance. Six of these items are from the E scale, with almost as many from the N and two from the L scale. Item content is consistent with a carefree extraversion. None of the interpretable first-order components loaded on S1.

S2. Salients -3, -4, -7, -8, -25, 35, -48, -51, -59, -63, 69, -82, -84, -88, and Desirability. Items -12, -44, -47, -63, -73, and 89 on Varimax only. Item 68 on Promax only. 12.3% of second-order variance, and 8.6% of reliable variance. Eleven of the 21 salient items are from the L scale, with the majority of the remainder being from the N scale. Most loaded negatively, with the sole positive loadings being the Desirability scale and Item 35 (doing as told as a child). Most of the item content relates to undesirable traits and activities, such as greed, slander, cheating, and irritability. It is clear that this is primarily a social desirability component. This is consistent with Eysenck's interpretation of the L scale and also consistent with the high correlation of the N items with Desirability, as reported in Table 3. It differs from the MMPI

L scale, which is implicated with a factor called defensiveness as well as with desirability (Jackson & Messick, 1962). It may be that the difference between social desirability value and *p* value, thought to underlie the defensiveness response style, is not sufficiently prominent in the case of the EPQ. Of the 9 interpretable first-order components 4 (Components IV, VI, XXII, and XXIII), loaded on this component. These components are primarily based on the L and N scales, one of them (VI), being based on E.

S3. Salient -55. Items 3, -15, -27, -34, 36, -42, -54, -58, -65, 82, -84, and 86 on Varimax only. 9.3% of second-order variance and 6.6% of reliable variance. The majority (7) of the salient items are from the N scale, with 4 of the remaining items from E. Item content is similar to factor S1, but the emphasis is more upon excitement and lack of worry per se, while S1 includes more items relating to people. None of the interpretable first-order components loaded on S3.

S4. Salients -14, 15, 27, -32, 45, 54, 62, 68, 77, and -Desirability. Items -2, -10, 12, -13, 21, 23, 31, 34, 38, -52, -60, 66, -70, 72, 80, and -82 also only on Varimax. 8.0% of second-order variance and 5.6% of reliable variance. Most of the salient items are from the N scale, with 9 others from the E scale. Three items (Items 12, 23, and 66) also loaded on Component XXII at the first-order level, and most of these items do match a general impression of neuroticism. Two interpretable first-order components (I and II) loaded S4. One of these is based on E-scale and one on N-scale items.

S5. Salients 5, -42. Items -13, 49, 54, 56, -65, 69, and -77 on Varimax only. Item -21 on Promax only. 6.9% of second-order variance and 4.8% of reliable variance. With the exception of Item 69 (dodging taxes), this component is clearly sociability, with 5 of 10 items being from the E scale. First-order Component XXIV loaded S5, which supports the sociability interpretation.

S6. Salients -10, -32, -64, and 68. Items -38, -47, and -65 only on Varimax. 5.5% of second-order variance and 3.8% of reliable variance. These items are a mix of the E and N

scales in equal proportions with one item from P. The item content and direction of loadings suggests introversion, combined with a denial of worry. The presence of Item 68 (wishing to be dead) is somewhat odd and consistent with depression. First-order Component VIII also loaded S6. This component is a mix of items from the L and P scales.

S7. Salients -40, -48, -52, and -66. Items -35, 41, 50, -61, -80, -81, and -88 only on Varimax. 4.8% of second-order variance and 3.3% of reliable variance. This component loads items from all four scales, four of the salient items being from the N scale. This component also hints at a quiet introvert, but there is a strong indication of a denial of problems. Component VII loaded S7. This was based on the P scales.

S8. Salients 26, 40, 46, 50, and 83. Items -11, 13 and -84 on Varimax only. Items 24 and -42 on Promax only. 3.8% of second-order variance and 2.7% of reliable variance. Five of the salient items are from the P scale, and much of the item content is consistent with such an interpretation, although this would necessitate a socially outgoing psychoticism, involving keeping promises (Item 14) and enjoying other people (Items 40 and -42). No interpretable first-order component loaded S8.

Among these eight second-order components, none load items from only one EPQ scale. Furthermore, only S5 and S7 load on clearly interpretable first-order components. Some components (S1, S3, S4, S6, S7, and S8) suggest aspects of the EPQ scales, but the interpretation is not clear and could easily be questioned. None of these latter components is dominated by items of the scales suggested by the content of the salient items, which in several cases are from different scales. In addition, the majority of the interpretable first-order components do not load on the second-order components in a manner consistent with Eysenck's claimed structure. It is noteworthy that the three first-order components based on E (II, VI, and XXIV) did not load on the same second-order component as Eysenck's model requires.

Third-Order Analysis

At this point, extraction of the structure suggested by Eysenck from the EPQ has been suggestive at best. However, a third-order (T) components analysis was conducted on the basis of the correlations of the oblique axes of the second-order Promax rotation. These correlations are reported in the upper triangle of Table 6. Varimax and Promax rotations were carried out as for the second-order components.

Four third-order components (eigenvalues of 1.91, 1.58, 1.15, and .91) accounted for 69.4% of the variance of the second-order correlation matrix (32.9% of the total variance). One item had a multiple loading on the Promax rotation, and 70 items failed to load on any of the four third-order components. On the Varimax rotation, 34 items failed to load on any component.

Two second-order components loaded each of the four third-order components, as well as a number of the original items. Only item loadings greater than $|.4|$ were interpreted. Items were salient on both rotations unless otherwise indicated.

T1. Salients S1 and S8. Items -11, -34, -84, and -88 on both rotations. Items -3, -7, -8, 10, 13, -19, 26, 32, 36, -47, -48, 56, -58, 64, -66, and -73 on Varimax only. 23.9% of third-order variance and 11.3% of reliable variance. Nine of the salient items are from the N scale. S1 is largely Extraversion and S8 contains Psychoticism content related to social activities.

T2. Salients S5 and S6. Item 16 on both rotations. Items 5, -42, -45, 49, -65, and -83 on Varimax only. 19.8% of third-order variance and 9.4% of reliable variance. The two salient second-order components suggest a social introversion interpretation. This is also found at the item level and is consistent with extraversion. Of the 7 salient items, 4 are from the E scale.

T3. Salients S3 and S4. Items 15, 27, 34, 38, 54, 62, 72, 77, -82, and -Desirability on both. Items -1, -2, 12, 21, 23, -24, 31, -40, 50, 75, 80, and -86 on Varimax only. 14.4% of third-order variance and 6.8% of reliable variance. Both second-order components and the majority (13 of 21) of the salient items are based on

Neuroticism. The salience of social desirability is noteworthy.

T4. Salients S2 and S7. Items -35, 54, 63, 69, -78, and -89 in both. Items 25, 44, and -52 on Varimax only. 11.3% of third-order variance and 5.4% of reliable variance. S2 is a mix of Lie and Neuroticism, while S7 is largely Neuroticism. However, the pattern of salient items strongly suggests the L scale, with 7 of 9 salient items from that scale.

In general, the results of the item factor analyses are only weakly supportive of the claimed structure of the EPQ. There are components at all three levels which are based upon items from the four Eysenck scales, but the claimed four-factor structure does not strongly present itself. Many items did not load on any component, and components were formed from items from more than one scale.

Orthogonality of Scales

Eysenck and Eysenck (1975) state the four EPQ scales are meant to be mutually independent. Table 7 presents both the correlations among the third-order components (upper triangle) and among the sets of scale scores (lower triangle). The significant negative correlation of the E and N scales is larger than those reported in the manual (Eysenck & Eysenck, 1975, Table 6), as is the significant correlation between the L and E scales. A signi-

ficant correlation between the L and P scales, as reported in the manual, was not found. These differences may reflect differences between Canadian and British respondents. In general, the claim of orthogonality of the four scales is supported.

Difficulty Factors

Factor analyses of binary response items frequently raise the accusation of difficulty factors. Gorsuch's example (1974, p. 261) of hypothetical difficulty factors and Ferguson's (1941) example of real difficulty factors are both based upon ability data. Examples of difficulty factors in personality data are rare, despite the much greater frequency of factor analysis of personality items than of ability items. Why are they not reported? One possibility is that no one looks for them.

Such a search was made in this case, using two methods. The first was based on the observation that in Gorsuch's (1974) example, p values correlate either $\pm.5$ or $\pm.95$ with the rotated factor loadings. This implies that there is a strong possibility that a difficulty factor could be identified by a correlation of near unity with the actual p values. In the case of the 91 EPQ components dealt with here, the highest such correlation was $-.732$, and only 9 were greater than $.1$. The highest correlating component was the third, which was not interpreted; and the items salient

Table 7
Correlations among Scales and among
Third-Order Components

	1	2	3	4
1	-	.03	-.06	-.04
2	-.06	-	-.05	-.30
3	.12	-.29**	-	.31
4	-.08	-.14*	-.27	-

Note: Decimals omitted. Variables are below the diagonal, third order components above the diagonal. Scales are P, E, N and L, from top to bottom. Components are in order of extraction.

* $p < .01$

** $p < .05$

on this component had neither similar item content nor similar p values.

The second method used the 13 items with extreme (less than .1 or greater than .9) p values as marker variables in a search among the 31 Varimax- and Promax-rotated components for components having these items as salients. The best component that could be located was Component XI, which had Items 48 and 61 salient. These items had p values of .93 and .97, respectively. This might appear to be a difficulty factor, except that other items with similar p values loaded at most .22 on this same factor. This does not appear to be a difficulty factor after all, then. It is perhaps safest to label it uninterpretable.

What procedure should then be adopted towards difficulty factors? McDonald and Ahlwat (1974) have presented theoretical reasons for discounting difficulty factors, and here an empirical search was made in a situation in which difficulty factors might be expected and were not located. It thus appears that the practical importance of difficulty factors for binary data derived from personality questionnaires is much less than has been assumed.

Discussion

There are two basic considerations regarding the properties of EPQ items to be discussed. The first of these concerns the basic psychometric properties, such as p values. The second concerns the factor structure claimed for the EPQ items.

On the psychometric level the EPQ scales appear sufficiently reliable, although the internal consistency of the P scale is substantially lower than that reported by the manual and, as always, somewhat higher consistencies would be preferred for the other three scales. These do, at least, replicate the claimed reliabilities. The E, N, and L scales have a moderate tendency toward extreme p values, but the P scale is dominated by such items. This scale is substantially poorer in this regard than the average CPI or MMPI scale (Helmes, Reed, & Jackson, 1977).

The P scale is also heavily affected by social desirability at the item level (Table 3). This may be inherent in the definition of Psychoticism as an undesirable characteristic, but assessment of undesirable traits independent of their desirability should be possible (Jackson, 1970). As one example, the N scale correlated even more highly with desirability than did the P scale, although P is presumably more undesirable than N. In any case in which there is a substantial relationship between a scale and social desirability, the interpretation of scores on that scale becomes subject to the confounded effect of social desirability. This makes the construct in question effectively useless. This is true of both N and P scales of the EPQ.

On the item component analyses level, the different methods of analysis gave somewhat equivocal results. Analysis of first-order components showed that in all rotations used, a majority of the EPQ items had their highest loading on their own scale. The P scale was consistently the poorest in this regard; and the N scale, the best. However, if more components were extracted and rotated, the three-factor structure claimed for the EPQ (P, E, and N) fell apart and did not reappear at either second- or third-order levels. At the third-order level, there was an indication of factors similar to N, E, and L; but the number of salient items was low, and the evidence weak. This implies that to some extent, the item structure that has been stated to exist for the EPQ is, instead, being imposed upon it and is not revealing itself in Eysenck's analyses.

There is also a secondary issue with regard to analytic procedures. The factor analytic procedures used in the development of the EPQ are described by Eysenck and Eysenck (1976). For the most part, they involved the Promax rotation of three first-order components. Third-order components were also extracted and compared to the first-order components, and coefficients of similarity were calculated in excess of .9. Eysenck and Eysenck (1976) claimed that the two procedures give highly similar results for their data. Cattell (1966) has argued this is not true in all cases. The data in this analysis are

consistent with both arguments. Table 4 shows that 14, 16, 22, and 18 items loaded highest on P, E, N, and L first-order components with a Promax rotation. On the third-order Promax rotation, the corresponding numbers were 9, 6, 7, and 11. The two procedures appear to give substantially different results. However, coefficients of congruence between the Promax rotations of four first-order components and the four third-order components were all in excess of .95. Thus, substantially different opinions would be formed about the similarity of first-order and third-order structure from these two methods. The first approach emphasizes the highest loading of each item; the second uses all loadings, weighting them all equally. Two structures that look quite different may have very high coefficients of congruence. This may account for some of the differences of opinion over the structure of Eysenck's scales.

Nevertheless, if the claim is to be accepted that the three content scales of the EPQ represent the major dimensions of personality, then these scales should appear in later analyses. Howarth's (1976; Howarth & Browne, 1972) analyses of the EPI noted that the scales were not univocal. That result is replicated here for the EPQ, and also by Loo (1979). Furthermore, the higher order analyses did not replicate those reported by Eysenck and Eysenck (1976). The closest approximation is the appearance of about half of the L scale at the third order, together with evidence of a sociability component (which is not the same as extraversion) and a suggestion of neuroticism.

The analysis reported here was based on smaller samples than that reported by Eysenck and Eysenck (1976) and did not separate the sexes. It was also based upon Canadian rather than British respondents. An analysis combining the sexes should not result in a different structure in itself. Eysenck and Eysenck (1976, p. 53) have stated that their rotations for the two sexes are highly similar. A smaller sample may result in some differences, as there is more sampling error. However, a strong structure should be apparent even in a small sample. Examina-

tion of the Eysencks' analyses (1976; Table 12) show that a large proportion of the loadings of the salient items were below .5. Fully 41 of 90 keyed items fell below the criterion of .4 used here. It therefore does not appear to have a strong factor structure that could be replicated easily.

Finally, it remains possible that there are distinct differences between British and Canadian respondents. However, the scale means and standard deviations were fairly similar, and the cultural differences are not great. These reasons do not appear sufficient to account for the failure of the EPQ structure to appear whenever more than three or four components are extracted. From the number of items with low loadings in the original analysis (Eysenck & Eysenck, 1976), it is arguable whether the EPQ originally had the structure claimed for it.

It thus appears that the factor structure claimed results from extracting too few factors and from interpreting low factor loadings. The EPQ scales are not univocal. Furthermore, the P and N scales are heavily contaminated by social desirability, and a clear interpretation of these scales in terms of the definitions of psychoticism and neuroticism is not possible. The psychometric properties of the P scale are particularly poor. These conclusions are based upon a small sample at one university. They do, however, argue that the EPQ is not an improvement upon the EPI, itself not a good instrument for measuring personality.

References

- Berg, I. A. (Ed.), *Response set in personality assessment*. Chicago: Aldine, 1967.
- Cattell, R. B. Higher order factor structures and reticular vs hierarchical formulae for their interpretation. In C. Banks & P. L. Broadhurst (Eds.), *Studies in psychology*. New York: Barnes & Noble, 1966.
- Edwards, A. L. *The measurement of personality traits by scales and inventories*. New York: Holt, Rinehart, & Winston, 1970.
- Eysenck, H. J. *The manual of the Maudsley Personality Inventory*. San Diego: Educational and Industrial Testing Service, 1962.

- Eysenck, H. J. Superfactors P, E, and N in a comprehensive factor space. *Multivariate Behavioral Research*, 1978, 13, 475-481.
- Eysenck, H. J., & Eysenck, S. B. G. *Manual for the Eysenck Personality Inventory*. San Diego, CA: Educational and Industrial Testing Service, 1968.
- Eysenck, H. J., & Eysenck, S. B. G. *Personality structure and measurement*. London: Routledge & Kegan Paul, 1969.
- Eysenck, H. J., & Eysenck, S. B. G. *Eysenck Personality Questionnaire manual*. San Diego, CA: Educational and Industrial Testing Service, 1975.
- Eysenck, H. J., & Eysenck, S. B. G. *Psychoticism as a dimension of personality*. London: Hodder & Stoughton, 1976.
- Farley, F. H. Social desirability, extraversion, and neuroticism: A learning analysis. *Journal of Psychology*, 1966, 64, 113-118.
- Ferguson, G. A. The factorial interpretation of test difficulty. *Psychometrika*, 1941, 6, 13-17.
- Gorsuch, R. L. *Factor analysis*. Philadelphia, PA: Saunders, 1974.
- Helmes, E., & Jackson, D. N. The item factor structure of the Personality Research Form. *Applied Psychological Measurement*, 1977, 1, 185-194.
- Helmes, E., Reed, P. L., & Jackson, D. N. Desirability and frequency scale values and endorsement proportions for items of Personality Research Form-E. *Psychological Reports*, 1977, 41, 435-444.
- Hendrickson, A. E., & White, P. O. PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, 1964, 17, 65-70.
- Howarth, E. A psychometric investigation of Eysenck's Personality Inventory. *Journal of Personality Assessment*, 1976, 40, 173-185.
- Howarth, E., & Browne, J. A. An item factor analysis of the Eysenck Personality Inventory. *British Journal of Social and Clinical Psychology*, 1972, 11, 162-174.
- Jackson, D. N. A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 2). New York: Academic Press, 1970.
- Jackson, D. N. *Personality Research Form manual* (Rev. ed). Port Huron, MI: Research Psychologists Press, 1974.
- Jackson, D. N., & Messick, S. Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology*, 1962, 65, 285-299.
- Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.
- Loo, R. A psychometric investigation of the Eysenck Personality Questionnaire. *Journal of Personality Assessment*, 1979, 43, 54-58.
- McDonald, R. P., & Ahlawat, K. S. Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 1974, 27, 82-99.
- Rogers, T. B. The process of responding to personality items: Some issues, a theory, and some research. *Multivariate Behavioral Research Monographs*, 1971, 6 (Monograph No. 71-2).
- Schönemann, P. H. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 1966, 31, 1-10.

Acknowledgments

I thank R. D. Chrisjohn for his collaboration in the collection of the data and S. Leboldus for her dedication to retyping tables. Funds for data analysis were provided through the Department of Psychology, The University of Western Ontario.

Author's Address

Send requests for reprints or further information to Edward Helmes, Department of Psychology, London Psychiatric Hospital, P.O. Box 2532, Terminal A, London, Ontario, Canada N6A 4H1.