

The Item Factor Structure of the Personality Research Form

Edward Helmes and Douglas N. Jackson
The University of Western Ontario

Multiscale personality inventories have rarely, if ever, demonstrated factor structure broadly consistent at the item level with scale keys. An item factor analysis of the 352 items of PRF-E was undertaken to evaluate the extent to which PRF items define separate and distinct factors corresponding to keyed scales. A matrix based on 352 items and 214 subjects drawn from 31 colleges was decomposed according to the Eckart-Young theorem, the factor loading matrix was formed, and then was rotated to an orthogonal target matrix of the 22 PRF-E scales. Inspection of the rotated matrix showed that only 2 of the 352 items failed to load in the keyed direction. The mean loading of items on their scale factor was .38; the mean loading of non-scale items was .09. This strong tendency for scale items to load more highly than non-scale items was also reflected in the majority of scale items being among the 16 highest loadings on each scale factor. For three scales, Dominance, Harmavoidance, and Order, the scale items were the 16 highest loaded items. The observed item factor structure of the PRF is interpreted as being attributable to the methods of construct definition and item selection used in the construction of the PRF.

Neither from the empirical evidence nor from the writings of some psychometric theorists might one reasonably expect clear-cut factorial structure based on an analysis of item data.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 1, No. 2 Spring 1977 pp. 185-194
© Copyright 1977 West Publishing Co.

Rarely have empirical data been presented which support the contention that factor analytically-based item structure corresponds to the keying of a published personality questionnaire. The presence of such an item structure would appear desirable, but it has seldom been made apparent. Most attempts at demonstrating such structure have used the 16 PF (Howarth & Browne, 1971; Eysenck & Eysenck, 1969; Karson & O'Dell, 1974) and none have been entirely successful. Cattell (1973) has criticized some of the reports on methodological grounds, but has failed to demonstrate superior structure with a new analysis based on his preferred methods. An attempt to do so by Burdsal & Vaughn (1974) does not demonstrate such a correspondence between factor loadings and scoring key. The data reported indicate a mean absolute loading of keyed items of only .25, and only 90 of 171 non-intelligence items load in the keyed direction.

A number of theoretical speculations have been advanced for the empirical difficulties in identifying a stable and consistent item structure. Items have been alleged to have a number of undesirable properties that make the identification of a consistent item structure more difficult. Items are said to: yield low reliability with corresponding low communality (Comrey, 1961; Cattell, 1973); require the use of product-mom-

ent correlations, like the phi coefficient, which were believed (erroneously) to yield difficulty factors (McDonald & Ahlawat, 1974); yield "bloated specifics" rather than generalizable factors (Cattell, 1973); be subject to cultural localisms in their interpretation; be saturated with response bias (Berg, 1967; Edwards, 1970); and depend in their factorial structure upon the sampling of subjects and of situations (Cattell, 1973). Although some of these problems may be not entirely without foundation, particularly those pertaining to the lack of high reliability and of the possible dependence of factorial structure upon certain respondent subgroup characteristics, there is merit in seeking evidence for an hypothesized factorial structure at the item level. Items form the building blocks of other measures; items that lack at least a moderately clear structure will provide an uncertain basis for developing scales with superior factorial qualities.

There have been few attempts to outline the pre-conditions in item development and selection which would permit the identification of a clear structure at the factorial level. It might be useful to identify some of these. First, there is now growing evidence that items should meet some formal or informal judgmental criteria regarding substantive and content validity; that is, there should be a clear and demonstrable relation between the item content and the construct which the item is intended to measure. There is evidence, for example, that groups of judges in multidimensional scaling and in social perception studies can agree about the relationships between item and scale content as demonstrated by reliabilities of above .98 (Boyd & Jackson, 1967; Jackson, 1970; Reed & Jackson, 1975; Verden, Jackson, & King, 1969). Other evidence suggests that even novice item writers will approach the task of scale preparation from the point of view of the scale definition and can develop scales with external validities greater than those developed using empirical methods of item selection with respect to a criterion group (Ash-ton & Goldberg, 1973; Jackson, 1975).

Second, items should be free from response bias. A large general factor has been frequently identified in personality measures associated with the judged desirability of items (Jackson & Messick, 1961; Edwards, 1970). While the role of evaluation in self-presentation is hardly surprising and indeed may be an important theoretical and empirical question, using items highly loaded on such a broad general factor places the investigator in an extremely weak position if his goal is to identify a large number of specific dimensions of personality. Specific substantive and empirical attention is required to suppress this general dimension if items are to be highly saturated with respect to specific dimensions.

Third, care must be exercised in item selection both substantively and empirically to avoid the intrusion of irrelevant content in the specification of an item designed to reflect univocally a particular construct. Probably the most sensitive means for avoiding such intrusion is to employ careful judgmental criteria. Empirical evidence of the univocal properties of items is also required.

Fourth, since at least two points in a multidimensional space are required to define a straight line, it is useful to have both ends of the dimension anchored respectively with positive and negative exemplars of the construct. Although this has been recognized as a desirable characteristic of attitude scales for more than forty years, there are nevertheless still personality questionnaires being published with grossly imbalanced keying.

Fifth, it is good practice to avoid high item redundancy and the "bloated specific" problem by judgmentally evaluating the results from any empirical item-analytic procedure for content generalizability. For example, we found while developing a Risk-Taking scale, that the best four items statistically all concerned gambling. Two of these were arbitrarily eliminated and replaced with other facets of risk-taking behavior.

The five procedures outlined above for enhancing a clear factorial structure at the item level do not necessarily require the actual com-

putation of an item factor matrix for empirical item selection. Indeed, the number of items required for the pools of major multiscale personality inventories would strain any but the very largest digital computers. It should be noted that even personality inventories claiming to be factor-analytically based are rarely constructed on the basis of a published item factor analysis. Furthermore, equivalent results are obtainable without massive item factor analyses, a fact not recognized by everyone. Most observers of personality assessment distinguish between factor-analytically based or derived personality inventories and other methods of scale construction.

Henrysson (1962) has demonstrated, however, that the point-biserial and biserial correlation coefficients can be interpreted as simple functions of the first centroid factor loading of a set of items. The point-biserial is simply the first centroid factor loading of the item divided by the item standard deviation, or, in the case of items with equal standard deviations, is a simple function of the first centroid loading. The biserial is the first centroid factor loading divided by the ordinate of the normal distribution corresponding to the proportion of individuals endorsing the item. Neill and Jackson (1970) found that there was a correlation above .99 between indices of item-scale association based upon point-biserial and biserial correlations and the first principal component factor loading.

The Personality Research Form, Form E (PRF-E), having undergone two separate sequential item analyses, in addition to having been developed in a manner consistent with the recommendations outlined above, might be expected to show some evidence of consistent factorial structure at the item level, even though it has not generally been regarded as a factor-analytically based personality questionnaire. It is thus the major aim of this study to evaluate the PRF in terms of its item factor structure to determine the extent to which scale factors emerge as a by-product of its development. This evaluation can perhaps best be done by an examination of the degree to which the PRF-E items correspond to the scoring key.

Procedure

Subjects

Subjects were 214 college students, 100 males and 114 females, volunteers drawn from randomly selected names from each of a random sample of 31 North American colleges covering a wide range of student populations and geographic regions. An official at each institution was requested to provide a student directory or other listing of students' names and addresses. For those colleges that replied and provided such a directory (five did not), a random sample of students was selected and addressed a written request for their cooperation in a research study. A covering letter together with the test materials and a return envelope were provided to those who replied favorably.

Materials

The details of PRF-E scale construction and interpretation are presented in the second edition of the PRF Manual (Jackson, 1974), but a brief outline might be helpful here. PRF-E uses the same 22 scales as the long forms AA and BB. It is based on the original PRF item pool and of the 352 PRF-E items; 14 are not found on either forms AA or BB. The wording of 89 of the remaining PRF-E items differs slightly from the AA-BB version.

The first step in the development of PRF-E was an explicit definition of the constructs to be used in the scales, incorporating both positive and negative aspects of the traits in question, and a faceted grid of situations and modes of expression. This was followed by item writing, which involved well over 100 items for each scale. The next step was an initial item analysis, in which desirability was minimized through the use of the Differential Reliability Index (Jackson, 1970). On the basis of this analysis, 880 items were assigned to two parallel forms containing 22 scales of 20 items each. These scales were re-administered to a new group of subjects, and a special algorithm was employed for minimizing redundancy among the items (Neill & Jackson, 1976). Using this information, the best

16 items for each of 22 scales were selected for PRF-E, and were further edited for simplicity of wording.

These steps were generally designed to foster discrimination at the item level, as well as at the scale level. Large item pools for each scale were designed to cover the entire range of possibilities as conceived by the definitions of each trait. A procedure was used which was designed to enhance the internal consistency of each scale by item analysis based on biserial correlations between the items and their appropriate scales. Suppression of desirability through the use of the Differential Reliability Index acted to prevent the appearance of desirability as the first factor, as it frequently does when this is not done. Minimum redundancy item analysis was designed primarily to foster discriminant properties of the scales.

Analysis

As a first step, the items were rearranged so that all items from a given scale were grouped together for convenience in further analysis. The row normalized data matrix of 352 items by 214 subjects was decomposed according to the Eckart-Young (1936) theorem using a singular value decomposition subroutine (Businger & Golub, 1969). The first 22 principal components were rotated using an orthogonal procrustes procedure (Schönemann, 1966). The target matrix of the procrustes procedure consisted of the PRF-E scoring key rearranged to correspond to the item order in the data matrix. Entries in the target matrix were +1 or -1 for true- and false-keyed items respectively, or 0 for non-keyed items. Each column of the target matrix was keyed to one of the 22 PRF-E scales. The Infrequency and Desirability scales were included in the analysis in order to determine the degree to which the control items load appropriately in comparison to the content items. It was also thought that the inclusion of these scales would serve to keep desirability and other relevant response sets separate from the substantive content of the other scales. Comparisons of interest in the resulting rotated factor loading

matrix are those between keyed and non-keyed items for each scale.

Following the completion of the orthogonal analysis, an oblique rotation (Horst, 1965, p. 394) was used to determine to what extent the oblique solution differed from the structure found in the orthogonal case.

Procrustes rotations have been criticized as allowing capitalization on chance within a matrix under certain conditions and as allowing the interpretation of spurious factors (Horn, 1967; Humphreys, Ilgen, McGrath, & Montanelli, 1969). These objections do not apply to the present analysis. Generally, at least four hypothesized loadings are considered sufficient to define a factor in a targetted rotation. The use of 16 targetted items per factor in the present study largely precluded arbitrariness in this regard.

Nevertheless, to evaluate the possibility that the results of the targetted rotation were due to capitalization on chance, the Jackson-Morf method (1974) of testing the null hypothesis was used. This method involves the formation of a randomly-derived target matrix which is uncorrelated with the rationally-constructed, non-random matrix, followed by orthogonal rotation to the random target. If there is no structure underlying the targetted rotation, then rotation to the random target should produce results equivalent to those found following rotation to the non-random target.

Results

Of the 352 items only two did not load in the keyed direction on their appropriate factor. The mean item loading for items on their own scale was .38; the mean absolute loading for items on scales other than their own was .09. The results are summarized by scales in Table 1.¹

¹Copies of the orthogonal and oblique rotated matrices, plus the table of correlations between factor scores and raw scores have been deposited with the National Auxiliary Publications Service. Order NAPS document No. 02856 for 36 pages of material. Order from ASIS/NAPS c/o Microfiche Publications, 440 Park Avenue South, New York, New York
(footnote continued on following page)

Table 1
Mean Absolute Factor Loading of Scale and
Non-Scale Items in Rotated Matrix

Scale	Mean Loading of Scale Items	Mean Loading of Non-Scale Items
Abasement	.289	.080
Achievement	.326	.079
Affiliation	.390	.092
Aggression	.347	.097
Autonomy	.402	.080
Change	.368	.117
Cognitive Structure	.406	.082
Defendence	.424	.078
Dominance	.543	.082
Endurance	.349	.076
Exhibition	.511	.085
Harmavoidance	.499	.074
Impulsivity	.407	.095
Nurturance	.416	.075
Order	.504	.070
Play	.258	.100
Sentience	.251	.088
Social Recognition	.286	.093
Succorance	.255	.083
Understanding	.317	.078
Infrequency	.502	.195
Desirability	.306	.077

Note. False-keyed items have been reflected.

Table 2 provides information on the rank-order structure of the factor loading matrix. In forming this table, items were rank-ordered in terms of the absolute value of the factor loading from highest to lowest, and the rank orders of the scale items tabulated. For three scales, Dominance, Harmavoidance, and Order, the scale items were the 16 highest-loading items in the factor loading matrix for those scales. Three scales had 15 scale items in the 16 highest-loading items, three scales had 14 items, and three

scales had 13 loading items in the first 16 items. The poorest scales in this sample had 8 scale items in the 16 first highest-loading items. The oblique solution gave roughly similar results, with factor loadings being somewhat higher overall (.42 for scale items and .13 for non-scale items). Only results from the orthogonal case are reported here because the orthogonal case is structurally more demanding than the oblique case.

Only 69.0% of the items loaded in the keyed direction for the random target matrix, as opposed to 99.4% loading in the keyed direction following rotation to the non-random target matrix ($p < .001$). In addition, the mean absolute loading for keyed items rotated to the random target was .24 compared to .38 for the non-random target.

(continued from previous page)

10016. Remit in advance for each NAPS accession number. Make checks payable to Microfiche Publications. Photocopies are \$9.00, microfiche are \$3.00. Outside of the U.S. and Canada, postage is \$2.00 for photocopy or \$1.00 for a fiche.

Table 2
 Rank Order Structure of the Rotated Matrix.
 Number of Scale Items, as Rank-Ordered Within Scales

Scale	Ranks 1-16	Ranks 17-48	Ranks 49-352
Abasement	10	3	3
Achievement	12	3	1
Affiliation	14	2	0
Aggression	13	3	0
Autonomy	13	3	0
Change	11	3	2
Cognitive Structure	14	1	1
Defendence	15	1	0
Dominance	16	0	0
Endurance	13	2	1
Exhibition	15	1	0
Harmavoidance	16	0	0
Impulsivity	15	1	0
Nurturance	14	2	0
Order	16	0	0
Play	8	2	6
Sentience	9	3	4
Social Recognition	11	2	3
Succorance	8	4	4
Understanding	10	4	2
Infrequency	6	6	4
Desirability	11	3	2

Slightly over one-half of the items in the random target matrix are keyed identically to items in the non-random target matrix. To overcome this overlap in the target matrices, a modified Jackson-Morff method was used, in which the formation of the random target matrix was completely random within the constraint that the number of positively and negatively keyed items match that in the non-random target matrix, i.e., 8 positively and 8 negatively keyed items per factor. Two such random targets were generated and orthogonal rotations to these targets were carried out in the same fashion as for the other rotations. The two random targets had 8 and 7 items keyed identically to the non-random target matrix and had 84.4% and 81.5% respectively of the rotated items loading in the same direction as keyed in the target matrix. These values are significantly lower than the 94.4% loading in the

keyed direction for rotation to the non-random target ($p < .001$ in both cases). The mean absolute loadings for keyed items in both rotated matrices was .16, as contrasted to .24 for the Jackson-Morff random matrix and .38 for the non-random, rationally-constructed target matrix. These results indicate that the pattern of substantial loadings following the rotation to the non-random target could not realistically be attributed to capitalization on chance, but reflect a solid item structure for the PRF-E.

The reliability of the analysis was assessed in two ways. First, internal consistency reliabilities of the 22 factors were assessed by the method described by Horn (1969). These are given in the first section of Table 3 and range from .98 to .68 from the first factor extracted to the twenty-second. Second, split-half reliabilities of the factors were calculated from a separate analysis

based on two sets of 176 PRF-E items each. Factor scores were calculated for the same 214 individuals on each set of 176 items and were inter-correlated. The split-half reliabilities based on the sets of factor scores are given in the second section of Table 3.

The correlations of the item factors for each scale with the raw scale scores are as follows (in alphabetical order of content scales, with Infrequency and Desirability last): .24, .30, .28, .30, .32, .19, .32, .19, .38, .30, .28, .41, .25, .29, .29,

.26, .36, .19, .29, .22, .16, .31. These values appear low, but there are several reasons why they might not be expected to be high. First, the maximum reliability is less than 1, due to the less than perfect reliabilities of the raw and factor scores. Second, the reliable variance of the test is dispersed among items and is less concentrated than it would be if factors based on scale scores rather than items had been used. Third, the factor scores are orthogonal while the raw scores are not. Oblique factor scores would cor-

Table 3
Item Factor Reliabilities

Internal Consistency		Split-Half	
Unrotated Factor	Reliability	Scale	Reliability
1	.98	AB	.22
2	.93	AC	.38
3	.92	AF	.42
4	.90	AG	.48
5	.88	AU	.28
6	.87	CH	.24
7	.85	CS	.44
8	.84	DE	.54
9	.83	DO	.68
10	.81	EN	.53
11	.79	EX	.59
12	.77	HA	.67
13	.75	IM	.52
14	.74	NO	.48
15	.73	OR	.76
16	.73	PL	.55
17	.72	SE	.52
18	.71	SR	.62
19	.71	SU	.54
20	.69	UN	.59
21	.69	IN	.21
22	.68	DY	.43

Note. The internal consistency factor reliabilities were calculated on the unrotated factors using Horn's (1969) formula; the split-half reliabilities were based on correlations of two separate sets of factor scores derived from the decomposition and rotating of two independent sets of 176 items (8 items per scale). The reliabilities were adjusted to the full-scale length of 16 items by use of the Spearman-Brown formula.

relate more highly. Fourth, since factor scores are based on the linear combination of the weighted contributions of every item to each factor, each set of factor scores is based upon 16 relevant items and 336 irrelevant items. The latter may contribute more noise than reliable variance. If factor scores were calculated using zeroes rather than the actual values for irrelevant items (Horn, 1965), the error component would probably be reduced.

Discussion

These results indicate a substantial degree of correspondence between the intended structure of the PRF and that actually found. It should be emphasized that *all* of the PRF-E items in the first and second item analysis correlated more highly with their own scale than with any of the other 19 irrelevant content scales, the Desirability scale, or the Infrequency scale. One can consider all of the PRF factors "confirmed" in the sense that from 8 to 16 items were among the 16 highest loading items on the appropriate factor and the great majority of items defined their own factor by virtue of loadings in excess of .30. These results are particularly encouraging in view of the number of dimensions which have been imposed on this space. Every one of the 352 items had 21 irrelevant bipolar dimensions with which to be contrasted. However this is not to deny that there are several problems in this type of analysis and that the structure is less than perfect. The two main sources of this difficulty lie in the sample and with the use of items for analyses of this type.

In addition to the normal random variations in sampling, there are several factors in relation to the sample used for this study that make it less than ideal for the purposes of the study. Although the sample is probably fairly representative of North American college student volunteers, this guarantees that there are quite strong differences among the members of the population sampled. Thus, not all members of the sample would exhibit the level of verbal ability and articulation of their self-concept that would

best reveal the item structure of the test in question. For example, average SAT verbal scores range from 364 to 700 among the colleges from which student directories were requested (Cass & Birnbaum, 1975). The subjects were volunteers and it cannot be said what effect this would have on the final results. In addition, the PRF-E was administered by mail, which means conditions of administration were not standardized. It would be of interest to compare the obtained factor structure with one derived from standard administration using a sample drawn from a college distinguished by students with high verbal ability. There is also, inevitably, a variation in the degree to which students exhibit response biases. This is compounded by the differential tendency of the test items to elicit such response biases. These have not been found to be major sources of variance on the PRF (Jackson & Lay, 1968) but will nevertheless exert some effect.

In examining the PRF-E, there is the relative unreliability of the items as compared with scale scores. This can be demonstrated by a comparison of the reliabilities of the item factors reported here and the scale reliabilities in the PRF Manual (Jackson, 1974). The latter are substantially higher than the former. In addition, in certain cases the definitions of scales are similar, even though they were designed to be non-overlapping. For example, it is difficult to discriminate a negatively-keyed Autonomy item from a positively-keyed Succorance item on the basis of content alone. In spite of what may appear to be extraordinary statistical efforts to minimize scale intercorrelations, there is nevertheless a tendency for such scales and their items to be correlated.

In spite of these problems these results have positive implications both for methods of test construction and for procedures of construct definition. Results of this analysis indicate that the techniques of scale construction already outlined can result in a test in which items reflect a factorial structure consistent with their keying. As far as we are aware, this relation has not been demonstrated for other methods of scale construction. In addition, the results indicate that

criticism of items as the basis for test construction may have been excessive. The five recommended steps in test construction can also be seen as properties necessary for optimal construct definition. A construct should exhibit both convergent and discriminant validity at the substantive level (Loevinger, 1957) among other things, and this can best be done by paying careful attention to developing these properties at the item level.

The crucial finding for this study is the confirmation at the factorial level of the 22 PRF-E scales, even though factor analysis was not explicitly employed in scale development. Further work designed to uncover the higher-order item structure as well as the role of differential respondent sampling appears to be clearly in order.

References

- Ashton, S. G., & Goldberg, L. R. In response to Jackson's challenge: The comparative validity of personality scales constructed by the external (empirical) strategy and scales developed intuitively by experts, novices, and laymen. *Journal of Research in Personality*, 1973, 7, 1-20.
- Berg, I. A. *Response set in personality assessment*. Chicago: Aldine, 1967.
- Boyd, J. E., & Jackson, D. N. The perceived structure of social attitudes and personality: A multidimensional scaling approach. *Multivariate Behavioral Research*, 1967, 2, 281-297.
- Burdsal, C. A., & Vaughn, D. S. A contrast of the personality structure of college students found in the questionnaire medium by items as compared to parcels. *Journal of Genetic Psychology*, 1974, 125, 219-224.
- Businger, P. A., & Golub, G. H. Algorithm 358. Singular value decomposition of a complex matrix. *Communications of the Association for Computing Machinery*, 1969, 12, 564-565.
- Cass, J., & Birnbaum, M. *Comparative guide to American colleges*. 7th edition. New York: Harper & Row, 1975.
- Cattell, R. B. *Personality and mood by questionnaire*. San Francisco: Jossey-Bass, 1973.
- Comrey, A. L. Factored homogeneous item dimensions in personality research. *Educational and Psychological Measurement*, 1961, 21, 417-431.
- Eckart, C., & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, 1, 211-218.
- Edwards, A. L. *The measurement of personality traits by scales and inventories*. New York: Holt, Rinehart & Winston, 1970.
- Eysenck, H. J., & Eysenck, S. B. G. *Personality structure and measurement*. London: Routledge and Kegan Paul, 1969.
- Henrysson, S. The relation between factor loadings and biserial correlations in item analysis. *Psychometrika*, 1962, 27, 419-424.
- Horn, J. L. An empirical comparison of methods for estimating factor scores. *Educational and Psychological Measurement*, 1965, 25, 313-322.
- Horn, J. L. On subjectivity in factor analysis. *Educational and Psychological Measurement*, 1967, 27, 811-820.
- Horn, J. L. On the internal consistency reliability of factors. *Multivariate Behavioral Research*, 1969, 4, 115-125.
- Horst, P. *Factor analysis of data matrices*. New York: Holt, Rinehart & Winston, 1965.
- Howarth, E., & Browne, J. A. An item-factor-analysis of the 16 PF. *Personality*, 1971, 2, 117-139.
- Humphreys, L. G., Ilgen, D., McGrath, D., & Montanelli, R. Capitalization on chance in rotation of factors. *Educational and Psychological Measurement*, 1969, 29, 259-271.
- 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*. Goshen, New York: Research Psychologists Press, 1974.
- Jackson, D. N. The relative validity of scales prepared by naive item writers and those based on empirical methods of personality scale construction. *Educational and Psychological Measurement*, 1975, 35, 361-370.
- Jackson, D. N., & Lay, C. H. Homogeneous dimensions of personality scale content. *Multivariate Behavioral Research*, 1968, 3, 321-337.
- Jackson, D. N., & Messick, S. Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement*, 1961, 21, 771-790.
- Jackson, D. N., & Morf, M. E. Testing the null hypothesis for rotation to a target. *Multivariate Behavioral Research*, 1974, 9, 303-310.
- Karson, S., & O'Dell, J. W. Is the 16 PF factorially valid? *Journal of Personality Assessment*, 1974, 38, 104-114.
- Loevinger, J. Objective tests as instruments of psychological theory. *Psychological Reports*, 1957, 3, 635-694.
- McDonald, R. P., & Ahlward, K. S. Difficulty factors in binary data. *British Journal of Mathematical*

- and Statistical Psychology*, 1974, 27, 82-99.
- Neill, J. A., & Jackson, D. N. An evaluation of item selection strategies in personality scale construction. *Educational and Psychological Measurement*, 1970, 30, 647-661.
- Neill, J. A., & Jackson, D. N. Minimum redundancy item analysis. *Educational and Psychological Measurement*, 1976, 36, 123-134.
- Reed, P. L., & Jackson, D. N. Clinical judgment of psychopathology: A model for inferential accuracy. *Journal of Abnormal Psychology*, 1975, 84, 475-482.
- Schönemann, P. H. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 1966, 31, 1-10.
- Verden, P., Jackson, D. N., & King, G. A. Popular

conceptions of the etiology of alcoholism: A multi-dimensional investigation. *Quarterly Journal of Studies on Alcohol*, 1969, 30, 78-92.

Acknowledgements

This research was supported by Canada Council Grant number S74-0495 to Douglas N. Jackson and Philip L. Reed. Parts of this paper were presented at the meetings of the Society of Multivariate Experimental Psychology, Gleneden Beach, Oregon, November, 1975.

Author's Address

Edward Helmes, Department of Psychology, The University of Western Ontario, London 72, Canada.