AN INVESTIGATION OF SCORES EARNED BY ADULTS ON THE RAVEN PROGRESSIVE MATRICES TEST

C. Neil Shaw
Florida State University

The purpose of this study was to investigate the Raven Progressive Matrices Test, 1938 edition (PM), in order to draw some inferences concerning its possible usefulness as a predictor variable in an adult population. In order for a test to be useful for predictive purposes it must possess certain intrinsic characteristics in relation to the group for which predictions are desired. Those intrinsic characteristics investigated in this study were: item difficulty, item correlation with total test score, test raw score distribution, and test-retest correlation with a one week delay.

A review of the literature indicated that few normative studies have been reported using the PM in an adult population. A study of the type reported here would be a necessary first step in evaluating the usefulness of the test for adults. In a study by Eysenk (2) involving a sample of 100 British Civil Defense workers a test-retest correlation of .87 was found with a four week delay between testings. In a study conducted by Foulds (3) involving the 1947 revision of the PM a mean score of 29.8 was found for technical and commercial workers. The 1947 revision consists of the same types of items as the 1938 edition with the only difference being in the difficulty and complexity of the items.

The PM is an untimed, nonverbal test consisting of a graded series of logically designed patterns. The test is divided into five sections of twelve problems each. The PM is intended to "assess a persons maximum capacity to form comparisons by reason and analogy." (4) The instrument, being limited to highly abstract material, is a nonverbal instrument which serves to measure Spearman's "G" factor (induction from relations) and is thought to be related more directly to "active" abstract intelligence and less to academic achievement, educational opportunity, or cultural background than the verbal measures of ability. (1)

The Sample

The population from which the sample for this study was drawn consisted of inmates of the Federal Correctional Institution at Tallahassee, Florida. The sample was restricted to subjects who had completed the eighth grade. The sample was selected on a volunteer basis, and although no controls were used to insure randomness

it is the opinion of the writer that the sample could be considered as a random sample of the inmates who met the criterion for inclusion in the sample. The sample consisted of fifty-six subjects. However, the test-retest reliability coefficient was based on a sample of thirty-seven subjects. The mean age of the subjects was thirty-four; the mean educational attainment was grade eleven. However, several of the subjects indicated that they had received high school equivalency diplomas so the level of formal schooling would be somewhat less than grade eleven.

The data were gathered in small group administrations with no time limits imposed. The test was administered according to the instructions in the manual with the exception that some of the wording was altered slightly as the test was developed with instructions for administration to British subjects.

Results

The mean score attained on the first administration was 44.2 with a standard deviation of 12.1. The maximum attainable score of 60 is 1.3 standard deviations above the sample mean. The range of scores was from 14 to 59 correct responses.

Table 1 indicates the difficulty index of each item and the biserial correlation of that item with the total test score. Flanagan's table of the product moment coefficients of correlation corresponding to given proportions of success in the upper and lower 27% of the distribution was used to estimate the correlation coefficients.

Table 1

Item Difficulty Indices and Biserial Correlations

Item Number	Set A D	Set A r	Set B D	Set B r	Set C D	Set C r	Set D D	Set D r	Set E D	Set E r
1	100	.00	100	.00	94	.53	93	.60	71	.84
2	100	.00	100	.00	89	.68	89	.68	77	.70
3	100	.00	93	.60	91	.60	89	.68	59	.87
4	100	.00	94	.53	80	.70	87	.72	64	.84
5	100	.00	85	.72	82	.77	89	.68	75	.78
6	100	.00	85	.72	78	.77	80	.77	59	.84
7	89	.64	78	.14	84	.74	71	.69	52	.72
8	94	.44	69	.84	71	.84	77	.74	41	.30
9	98	.33	75	.90	69	.56	77	.77	37	.73
10	91	.54	75	.87	61	.82	66	.77	16	.74
11	80	.74	68	.87	48	.78	32	.51	9	.24
12	61	.53	57	.73	28	.56	30	.45	14	.68

Inspection of the table would indicate that all items except the last one in Set C, the last two in Set D, and the last five in Set E are not of sufficient difficulty to contribute significantly to the variance of the scores. Further inspection of the table indicates that with the exception of the extremely easy items the correlation of the items with total test score was high and positive.

The test-retest correlation for the thirty-seven subjects for whom data was available was .87. The mean score achieved on the first testing was 45.4 with a standard deviation of 9.6; the mean on the second testing was 47.7 with a standard deviation of 8.1. The difference between the means is not significant at the .05 level of significance. Due to the fact that the difference between the means is not significant and that there are some subjects who gained and some who lost from the first testing to the second it would not be accurate to state that the mean increase can be attributed to practice affect. However, the high gains for persons who initially had low scores may explain mean gain and may be an indication of practice affect only among those subjects who earned low scores on the initial testing. Table 2 shows the gain and loss in scores for the 37 subjects who were tested twice.

Table 2

Gains and Losses in Scores Between the Two Test Administrations

Initial Score	Second Test Score										
	Gain							Loss			
	11-12	9-10	7-8	5-6	3-4	1-2	0	1-2	3-4	5-6	7-8	9-10
58-60							1					
55-57						2	1					1
52-54				1	1	5						
49-51					2	1	1	1				
46-48						1		1		1		
43-45		1		1	1	1		1	1			
40-42	1			1					1			
37-39								1				
34-36	1				1							
31-33				1								
28-30					1							
25-27						1						
22-24	1											
19-21	1											

Caution should be exercised in interpreting these results as they do not represent the full range of scores attained on the initial testing. Inspection of the means would indicate that those scoring higher on the first testing returned for the second testing more frequently than those attaining low scores.

Summary

The results of the present study may be summarized as follows:

1. The individual items are not of sufficient difficulty to discriminate among the adult, male inmates in the sample.
2. Item correlations with the total test score are high when the difficulty level of each item is considered.
3. The test score distribution is negatively skewed indicating that the test does not discriminate at the higher ability levels.

The results of this study would indicate that the Raven Progressive Matrices,1938, does not possess the intrinsic characteristics which would make it a useful tool for most predictive purposes involving adult male subjects. Since this sample cannot be considered as a sample of the general adult population generalizations must be made with caution. However, it is the writers opinion that the test would not possess better intrinsic characteristics with other adult samples. This opinion is based on the fact that the Raven proved to be too easy for the prison sample. This criticism would most likely be even more severe for the general male population.

References

1. Cronbach, Lee J. Essentials of Psychological Testing. New York: Harper and Row, 1960, pp. 215-218.

2. Eysenk, H. J. The Effect of Incentives on Neurotics and the Variability of Neurotics as Compared with Normals. British Journal of Medical Psychology, XX (1951), 100-110.

3. Foulds, G. A. Variations in the Intellectual Activities of Adults. American Journal of Psychology, LXII (April, 1949), 238-246.

4. Raven, J. C. Guide to the Standard Progressive Matrices. London: H. K. Lewis and Co., 1960.