SOCIAL BEHAVIOR AND PERSONALITY, 1979, 7(2), 157-158

DOI 10.2224/sbp.1979.7.2.157

ESTIMATES OF FAKEABILITY ON THE EYSENCK PERSONALITY QUESTIONNAIRE

ROBERT LOO AND PAM WUDEL
University of Calgary

The fakeability of scales on the Eysenck Personality Questionnaire was estimated in a homogeneous sample of female undergraduates. Three instruction conditions were used: standard or honest, fake good, and fake bad. Various estimates of fakeability were used and the results supported Gordon and Gross's (1978) position that at lease 2 statistics be used so as to estimate both parameters, σ_{D2}, μ_{D2}, of the discrepancy distribution or of faking. The questionnaire appeared to be moderately susceptible to faking, at least in the fake good condition; however, the use of cutoffs on the lie scale should be considered to discard respondents.

Keywords: estimates of fakeability, Eysenck Personality Questionnaire scales.

The fakeability of self-report instruments – that is, the deliberate and systematic distortion of responses so as to create a particular impression – has been of concern to test constructors, researchers, and clinicians. Gordon and Gross (1978) have critiqued methods for operationalizing the concept of fakeability. In their approach to fakeability they stressed the utility of Naylor's (1967) index of accuracy, $\sigma_{e2} = \sigma_{D2} + \mu_{D2}$, where σ_{D2} is the variance of the distribution of the difference scores and μ_{D2} is the squared mean of the distribution of the difference scores. Naylor's index, unlike other indices of accuracy or fakeability, takes into consideration three rather than the usual one type of error. Two types of error – systematic error such as individual differences and random error such as error which affects the reliability of an instrument – affect σ_{D2}, whereas average or constant error such as that introduced by the situation affects μ_{D2}. Gordon and Gross (1978) noted that one popular index of fakeability, the mean difference ($mean_F$ - $mean_H$) between test scores obtained under honest (H) and faked (F) conditions, provides only an estimate of the magnitude of constant error, that is, μ_{D2}. On the other hand, the correlation between scores obtained under two conditions, r_{FH} is related to σ_{D2}. S_{D2}, the variance of the distribution of ($mean_F$ - $mean_H$) is an unbiased estimate of σ_{D2}, and, unlike r_{FH}, S_{D2} is sensitive to errors associated with variances of differing magnitudes in the distribution of honest and faked scores. Gordon and Gross (1978) recommended that at least two statistics, ($mean_F$ - $mean_H$) and S_{D2}, be used to estimate the parameters μ_{D2} and σ_{D2}.

In addition to the statistics discussed by Gordon and Gross (1978), some researchers have considered internal consistency coefficients to be useful indices of fakeability. Kirton (1977) divided a sample into high lie scorers and others based upon their scores on the lie (L) scale in the Eysenck Personality Inventory. He found that the group of high L scores had scores on five "Adorno" type the other group. From the assumption that it is easier to be consistent when tests which were slightly

Correspondence and reprint requests should be addressed to: Robert Loo, Bell-Northern Research, PO Box 3511, Station C., Ottawa, Canada KIY4H7.

more internally consistent than were the scores from telling the truth than when deliberately lying, he concluded that the comparable internal reliabilities supported the notion that high L scorers may include those who are both naive and honest. Farley and Goh (1976) computed internal reliabilities on scores from the psychoticism (P), extraversion (E), neuroticism (N), and L scales of the Eysencks' PEN. Reliabilities were obtained under three conditions: standard instructions and best and worst impression instructions. The reliabilities of the scales were generally as high or higher in the two faked conditions, best and worst impression, as in the normal or honest condition, and this was especially so for the P scale.

The purpose in the present study was to examine the effects of three instruction conditions, standard, fake good, and fake bad, on scores from the four scales, P, E, N, and L, in the Eysencks' latest inventory the Eysenck Personality Questionnaire (Eysenck and Eysenck, 1975). The effects of instructions were examined through the statistics ($mean_H$ - $mean_F$) S_{D2}, σ_{e2}, and r_{FH} as discussed by Gordon and Gross (1978), and through internal consistency reliabilities as measured using Cronbach's alpha coefficient.

METHOD

Participants And Procedure

Participants were 68 female undergraduates ranging in age from 18 to 21. At the first session the Eysenck Personality Questionnaire was administered under standard instructions in a group session, and eight weeks later the questionnaire was administered in a group a second time. However, no participant was aware that it would have been administered more than once. Half the questionnaires contained fake good and half fake bad instructions where the order of instructions was randomized. The instructions were identical to those used by Farley and Goh (1976) for the best and worst impressions, respectively.

Analyses

T tests were performed on the means of scores from the four scales which were administered under standard instructions to test for initial differences between the two groups, fake good and fake bad. Within each group, correlated tests were performed on the means of scores from the four scales which were administered under the standard or honest instructions and the instructions to fake responses to assess the influence of constant error, μ_{D2}, in this study the situation or instruction conditions. The Pearson correlation coefficients between scores in the honest and faked conditions, r_{FH}, were computed for each of the four scales. S_{D2}, the variance of the distribution of ($mean_H$ - $mean_F$), was also computed for the four scales. Naylor's (1967) σ_{e2}, the variance of the difference scores around the point of zero error, was computed for each scale. Finally, the internal consistency reliability as measured by the standardized alpha coefficient was computed for each scale under each instruction condition.

RESULTS AND DISCUSSION

Thirteen participants did not complete the questionnaires at both sessions; therefore the analyses are based upon 55 complete sets of data. The only significant t test on initial

group differences showed that the fake good group had a lower L score than did the fake bad group ($t = -2.04$, $df = 53$, $p < .05$). The results of the other analyses are presented in Table 1.

There were significant effects for the two faking conditions or situations, that is, for constant errors associated with μ_D on the P. E, N, and L scales as shown by the significant t tests. The one nonsignificant t test involved the L scale in the fake bad condition. Clearly, the large variance in the faking condition would prohibit the detection of a constant effect or mean difference.

Two extreme values were obtained for the statistic S_{D2}. A relatively small value of S_{D2} was obtained for the fake good condition on the P scale (2.19), thus indicating that the scale under this condition is not susceptible to errors associated with σ_{D2}. At the other extreme a very large value of S_{D2} was obtained for the fake bad condition on the L scale (428.48). These findings support Gordon and Goss's (1978) position that more than one statistic of fakeability is essential, and that the effects due to both μ_D and σ_{D2} must be estimated. Clearly, in the fake bad condition the L scale showed no effect for the parameter μ_D as indicated by the nonsignificant *t* value. However, there was a large effect for the parameter σ_{D2} as estimated by S_{D2}, a statistic which is sensitive to errors associated with variances of differing magnitudes. The large variance associated with the fake bad condition clearly indicates confusion about which direction would indicate faking "bad"; that is, it indicates individual differences in expectations about the appropriate direction for faking "bad".

TABLE 1

DISTRIBUTION STATISTICS INTERNAL CONSISTENCY, AND FAKEABILITY ESTIMATES OBTAINED UNDER THREE INSTRUCTION CONDITIONS ON THE EYSENCK PERSONALITY QUESTIONNAIRE

	Fake Good n = 32				Fake Bad n = 32			
	P	E	N	L	P	E	N	L
Honest								
mean	1.72	15.16	11.59	6.03	1.60	13.74	11.22	8.13
SD	1.42	4.10	5.03	3.63	1.30	4.41	4.62	4.00
α	0.34	0.83	0.85	0.77	0.28	0.84	0.82	0.78
Fake								
mean	1.09	18.16	4.96	17.34	21.86	4.60	15.39	6.00
SD	1.51	2.64	4.06	6.70	4.89	4.73	4.16	20.11
Alpha	0.58	0.77	0.84	0.92	0.96	0.92	0.86	0.92
r_{FH}	0.50	0.13	0.26	0.15	0.16	0.13	-0.32	-0.04
t test FH	-2.43	3.68	-6.53	9.53	20.06	-7.26	2.80	-0.49
S_{D2}	2.19	20.87	30.06	50.93	23.47	36.39	51.06	428.48
σe^2	2.44	29.03	70.56	177.31	432.96	118.17	66.26	414.39

Relatively moderate values of S_{D2} were obtained for the E and N scales, but the fake bad condition produced slightly higher values than the fake good condition. It seems, then, that these two scales are slightly more susceptible to faking "bad" in terms of GD^2 errors as estimated by S_{D2}. Similarly, the P and L scales are more susceptible to faking "bad" than "good" as indicated by values of S_{D2}.

The values of σ_{e2} appear to take into account both the effects of constant or situation error, as indicated by the *t* values, and of individual differences and random

error as indicated by the S_{D2} values. For example, the relatively small t and S_{D2} values obtained from the P scale in the fake good condition are reflected in the small value of σ_{e2} (2.44). In contrast, the larger *t* and moderate S_{D2} values from the P scale in the fake bad condition are reflected in the larger value of σ_{e2} (432.96), and the small *t* and large S_{D2} values from the L scale in the fake bad condition are reflected in the large value of σ_{e2} (414.39).

The high internal-consistency reliabilities of the scales under the two faking instructions in the present study and those obtained by Farley and Goh (1976) clearly show that, contrary to Kirton (1978), high internal consistency is possible when faking responses even on a multidimensional inventory of 90 items.

Finally, the faking instructions show, as expected, that the L scale is most sensitive to faking and that the other scales appear to be only moderately affected by instructions to fake good as compared with instructions to fake bad. If it is assumed that faking bad is not likely when this questionnaire is used, then the questionnaire may be judged to be moderately affected by faking. However, the cautious investigator may use a cutoff on the L scale to discard respondents.

One may consider the mean value in the fake good condition on the L scale (17) as the cutoff or the value two standard deviations above the mean under standard or honest instructions.

REFERENCES

Eysenck, H. J., & Eysenck, S. B. G. (1975). Manual of the Eysenck Personality Questionnaire. San Diego: Educational and Industrial Testing Service.

Farley, F. H., & Goh, D. S. (1976). PENmanship: Faking the P-E-N. British Journal of Social and Clinical Psychology, 15, 139-148.

Gordon, M. E., & Gross, R. H. (1978). A critique of methods for operationalizing the concept of fakeability. Educational and Psychological Measurement. 38, 721-782.

Kirton, M. (1977). Characteristics of high lie scorers. Psychological Reports, 40, 279-280.

Naylor, J. C. (1967). Some comments on the accuracy and the validity of a cu e variable. Journal of Mathematical Psychology, 4, 154-161.