

# Faking Personality Questionnaires: Fabricating Different Profiles for Different Purposes

ADRIAN FURNHAM  
*University College London*

Over fifty subjects completed four personality measures used in personnel selection. Based on a latin-square design they were asked to fill them in as they would if they were trying to present themselves as ideal candidates for the job of librarian, advertising executive or banker, while on one questionnaire they gave "honest," actual responses. The results indicated, as previous research in this area has shown, that the questionnaires are all highly susceptible to faking, and that a quite different prototypic profile arose for each of the three different jobs. The results are discussed in terms of the templates of fakers for specific jobs; the methods of detecting fakers, and what dissimulation studies tell us about theories of both occupations and personality.

Can job interview candidates fake on personality questionnaires used as selection devices? Can they produce different profiles depending on the job they are applying for? These are important questions given the growing use of such measures in industry. Robertson and Makin (1986) found about a third (35.8%) of 108 British organizations used personality tests in some selection decisions while just under a third (29.1%) used cognitive tests.

Reactions to "objective" psychological assessment by use of psychometric tests has been varied, ranging from cynicism and skepticism to euphoric, all-embracing enthusiasm. Many managers and personnel officers question the reliability and validity of personality tests, particularly with regard to the predictive and construct validity, the purpose for which they are most often sought. The most common of many objections to the use of personality inventories is that people lie or deceive cleverly on all self-report measures, especially personality questionnaires. Thus they are invalid because they do not yield true scores, especially on the assessment of all-important undesirable traits or behavior patterns. If the test relies on face validity, then this sort of faking will be a serious problem. If, however, criticism and construct validity have been established, then faking may have been (in some sense) taken care of. For example, if it has been shown that the best used-car salesmen are those who give a wildly exaggerated version of their own desirable personality traits, then such a profile, false though it is, will be a valid predictor of success at the job.

Faking can take one of three forms: people *deliberately sabotage* results by random, "patterned" or meaningless responses; there is *motivational distortion* or faking to achieve a particular profile (positive, desirable, employable or "faking bad"); and *sheer ignorance*, whereby the respondent, through lack of self-insight or self-under-

standing, cannot, rather than will not, accurately report on his/her, attitude, beliefs or behaviors. It is nearly always the second form of faking that is thought of as the most frequently occurring and damning. Academic discussion on these issues has revolved around the controversy concerning whether the term "faking" implies conscious vs. unconscious efforts to distort response patterns. (Archer, Gordon, Kirchner, 1987; Furnham, 1986a; Meredith, 1968). Lay people on the other hand use the term "fake" to imply a conscious effort at distortion which would not necessarily be related to unconscious efforts at test distortion, and this study concentrates exclusively on conscious efforts to dissimulate a particular profile.

Furnham (1986a) has recently reviewed the literature on response bias, social desirability and dissimulation. Various studies on the fakeability of personality questionnaires were critically reviewed. It has been demonstrated that questionnaires like the Eysenck Personality Questionnaire, the General Health Questionnaire, the Adorno F Scale, and the Vocational Interest Blank are all highly susceptible to faking and social desirability effects, so much so that highly desirable vs. undesirable profiles are relatively easy to produce. However some questionnaires, such as the Locus of Control Scale and the Self-Monitoring Scale, are far less susceptible to any kind of faking (Furnham and Henderson, 1982). It was concluded that questionnaires with low face validity and/or those measuring traits or behavior patterns not well understood or popularly known by the general public are less fakeable (Furnham, 1988). On the other hand some multidimensional traits that are highly fakeable, such as the Type A behavior pattern, show a particularly interesting pattern because positive traits like ambitious and achievement oriented are over-reported while negative behaviors like aggressiveness and freneticism are under-reported, so leaving the final score unchanged (Furnham, 1986b).

Furnham (1986a) drew three major conclusions from his reviews. The first was that no matter who the subjects were (students, policemen, army recruits, employed people) they showed a similar pattern of faking. Secondly, most studies have simply got subjects to fake cleverly or poorly but that some have got them to fake according to other instructions (i.e., fake a librarian or fake a mental patient) though the results have usually been predictable. Thirdly, studies done in real life, as opposed to experimental settings have yielded comparable results. This suggests that experimental work using *any* population group, and using "fake good" and "bad" instructions would yield comparatively robust results and replicable faking templates.

There has been a fair amount of research on the fakeability of self-report, personality measures used specifically in organizational settings. Furnham and Craig (1987) examined the fakeability of the Perception and Preference Inventory (Kostick 1977) by getting subjects to either fake good, fake bad or give an honest response. Four of the seven subscale scores, particularly work direction (indicating how hard a person works and how ambitious he or she is), leadership (preferences for, and ability in, the leadership role), activity (speed of work) and social nature (sociability), showed significant and predictable differences, with the fake good scoring significantly higher

than the fake bad group. As many organizations use personality questionnaires in their selection process it is of course a major concern for them to ensure that they are not getting badly distorted data. Recent studies of faked response-set characteristics in applied settings shows that it is relatively easy to fake some of the most widely used and respected measures (Archer et al, 1987). In a study of actual applicants to the British Metropolitan Police Force, Burbeck and Furnham (1984) found applicants had identical profiles to those student subjects who had been asked to fake skillfully on the questionnaire. The desire to be accepted as police recruits has led them to fake a socially desirable response pattern which in this instance meant high on extraversion and low on neuroticism. This conclusion is unwarranted in the absence of independent evidence and needs to be established. More recently, Furnham (1990) set out to examine the fakeability of three questionnaires used for personality assessment in applied and occupational settings: namely the 16 PF, the Myers-Briggs and the FIRO-8, which are extensively used in occupational selection and assessment in Great Britain. All the tests proved extremely sensitive to faking, especially the 16 PF. Furnham (1989) concluded from the results thus:

First in many instances the means and profiles of the two faking groups were not totally extreme showing that extremity is not necessarily an indicator of faking; second whereas some results may have been predicted or anticipated, others which may have yielded significant difference did not; third that not all findings were linear and often too much or too little of a quality or trait was considered equally good or bad; fourth not all the significant differences lay between fake good and control; finally that faking is not easy and there is not clear agreement as to what constitutes a good or bad response. (P ×)

Nearly all the faking studies have required subjects to "fake good" to examine the socially desirable profile. However, this method may have relatively low ecological validity because faking good in one context may not be equivalent to faking good in another. For instance if a personality test were given at an interview for a job as an actor the fake good profile would presumably be quite different from that of the ideal candidate aiming to join the army as an officer-cadet. Few studies have required subjects to fake particular profiles. Velicer and Weiner (1975) is an exception however, as their study required subjects to fake salesmen, fake librarian and fake ideal self. They found, as predicted, large differences between the resultant profiles, suggesting that subjects can fake many types of ideal or good professional profiles.

This study had a number of specific aims. First it set out to establish to what extent four, often-used personality tests were susceptible to deliberate faking. Comparisons were made between honest, control responses and those faking as if they were applying for one of three jobs: advertising executive, banker, or librarian. Secondly it sought to examine the nature of the typical fake profile in order to establish what traits or behavior patterns a subject group thought most (and least) desirable for each of the

three occupations. Thirdly the study attempted to identify the typical fake response pattern, which may be used to identify fakers.

## METHOD

### *Subjects*

Fifty-five subjects took part in this study; 20 were male and 35 female. They ranged in age from 18 to 41 years, though the majority were in their twenties. They were recruited from a British student population at London University. They were not paid for their participation but were given the results of the questionnaire where they responded honestly.

### *Questionnaires*

Each subject completed four questionnaires in random order such that there was equal chance that a test was done first, second, third or fourth.

1. *The FIRO-B* (Schutz 1978)

The Fundamental Interpersonal Relations Orientation-Behavior is a 54-item scale and each question is completed on a six-point response scale. The test measures three dimensions of interpersonal relationships: *inclusion* or the degree to which a person associates with others, *control* or the extent to which a person assumes responsibility, and *affection* or the degree to which a person becomes emotionally involved with others. The test is recommended for any sort of counseling and has limited, but satisfactory, psychometric properties in terms of validity and reliability.

2. *Vocational Preference* (Holland, 1985)

This is a 160-item/yes-no inventory designed to help people make vocational choices. While the original version offered scores on six dimensions this version provided eleven scores per subject. The measure is based on Holland's (1973) theory of vocational choice which has attracted most interest in the vocational guidance and occupational literature. The test has proven reliability and validity.

3. *Myers-Briggs Indicator* (Briggs, Myers and McCaulley, 1985)

This is a 166-item scale (Form F) which in most, but not all cases is based on a two-fold forced choice scale. It yields eight scores per person which results in a sixteen (2 x 2 x 2 x 2)-item taxonomic structure. The measure is based on Jung's type theory and was first developed over 40 years ago. It has impressive norms and satisfactory reliability and validity statistics. It is most extensively used in psychotherapy, career counseling and education.

4. *Kirton Adaptation-Innovation Inventory* (Kirton, 1976)

This is a 33-item questionnaire designed to locate a respondent on a single dimension which indicates the style of creativity characteristic of the individ-

ual. The inventory has attracted a fair amount of attention in the occupational literature and has shown itself to be both valid and reliable.

### Procedure

Subjects were randomly assigned to one of four conditions, A, B, C or D. The four conditions were given different instructions, namely:

**Fake Librarian:** "When completing this questionnaire we would like you to answer the questions as if you are applying for the job of librarian and attempting to present yourself in the best possible way for that job. You need not be honest in your answers."

**Fake Banker:** "When completing this questionnaire we would like you to answer the questions as if you are applying for the job of banker and attempting to present yourself in the best possible way for that job. You need not be honest in your answers."

**Fake Advertising Executive:** "When completing the questionnaire we would like you to answer the questions as if you are applying for the job of advertising executive and attempting to present yourself in the best possible way for that job. You need not be honest in your answers." **Control:** "When completing this questionnaire we would like you to be as honest as possible; that is, present yourself as you really are."

These instructions were given verbally to the subjects and repeated several times. The instructions are obviously quite different from those given when the tests are used in selection. The obvious, but implausible objection is that faking would occur only under such experimental instruction. Yet it is known faking occurs under standard nonfaking instructions. A full latin-square design was shown to the groups which indicated which of the four questionnaires they would complete honestly, and which they would fake. The design looked like this:

	Fake Librarian	Fake Advertiser	Fake Banker	Control
1. FIRO-B	A	B	C	D
2. KIRTON A-I	B	A	D	C
3. HOLLAND VPI	C	D	A	B
4. MYERS-BRIGGS	D	C	B	A

Hence each subject completed each of the four questionnaires and knew which to fake according to the instructions and which to respond to honestly. It took approximately two hours to complete the experiment and subjects were later given feedback on their individual and group scores.

## RESULTS

The results of each questionnaire were treated to a MANOVA to examine overall effects of response sets across subscales of each instrument followed by a one-way ANOVA per measured dimension followed by post-hoc comparisons (Scheffe tests) designed to reveal all differences at  $p < .01$ .

*FIRO-B*

The MANOVA was highly significant ( $F = 8.98, P < .001$ ) as were all the ANOVAs. The post-hoc comparisons give the clearest picture of the pattern of results. All comparisons between fake librarian and fake advertising executive were significant, indicating that the latter expressed and wanted more affection and inclusion but though they wanted more control, they expressed a lower need. There were only two significant differences between the faked advertising executive profile and the banker profile, both on the dimension of affection, and one between librarian and banker. Interestingly though there were few differences between the control group and the faked advertising executive and banker profiles, two major differences emerged on the honest self vs. fake librarian score, both with regard to honest control. Clearly the subjects perceived the ideal profile of a librarian to be radically different from that of an advertising executive, though the control group of honest responses was not much different from those of fake banker or advertiser.

*VPI*

Again the MANOVA was significant ( $F = 4.90, p < .01$ ), yet only just over half of the eleven dimensions yielded significant effects. On this scale, however, the major differences occurred not between the fake librarian and advertising executive as in the previous scale but between fake advertising executive and fake banker. The fake advertising executive had the highest artistic, but lowest conventional self-control, and masculinity score, while the fake banker had the lowest artistic, but highest conventional and masculinity score. Overall the subjects appeared to rate realistic as low and entrepreneurial as high for all three faked profiles and control, while artistic and conventional yielded the biggest differences.

*Myers-Briggs*

The MANOVA was highly significant ( $F = 14.11, p < .001$ ) as were all the ANOVAs. The size of the  $F$  levels and the number of significant post hoc differences indicates that of all the measures used in this study the Myers-Briggs is most susceptible to faking. Advertisers were perceived as having highest extraversion, lowest introversion, lowest sensing, highest intuiting, highest feeling and highest judging scores, while librarians had almost exactly the opposite—lowest intuiting and highest perceiving scores. The fake banker profile had the highest thinking and lowest feeling score and the highest perceiving score. The thinking-feeling dimensions yielded the largest and most dramatic differences.

*KAI*

The significant ANOVA showed substantial differences between the four groups' scores. Though the subjects tended to rate themselves as innovators they tended to

TABLE 1

	Means				ANOVA F	Post Hoc Comparisons					
	Fake Librarian	Fake Advertiser	Fake Banker	Control		FL	FL	FL	FA	FA	FB
	FL	FA	FB	C	Levels	FA	FB	C	FB	C	C
<b>FIRO B</b>											
Wanted Affection	2.26	6.41	2.76	5.06	9.37***	*			*		
Expressed Affection	1.73	6.00	1.92	3.86	8.33***	*			*		
Wanted Control	2.26	5.83	7.38	2.66	13.03***	*		*			*
Expressed Control	5.46	1.00	1.15	1.93	10.65***	*	*	*			
Wanted Inclusion	1.06	6.25	3.00	3.46	7.74***	*					
Expressed Inclusion	2.13	6.33	3.84	3.93	8.24***	*					
<b>VPI</b>											
Realistic	2.00	2.41	1.61	1.73	0.88						
Investigative	4.40	5.50	5.46	4.13	0.48						
Artistic	8.69	11.86	3.93	7.75	16.30***		*		*		
Scientific	5.30	4.413	4.13	10.00	0.26						
Entrepreneurial	9.33	8.66	8.23	7.86	0.14						
Conventional	5.69	1.73	9.33	1.08	18.33***				*		*
Self-control	13.38	8.33	10.33	9.25	5.36**	*					
Masculinity	6.07	5.13	8.73	6.66	7.66***				*		*
Status	8.30	8.46	11.26	9.66	5.87***		*		*		*
Infrequency	5.16	7.13	8.92	4.86	5.31**	*					
Acquiescence	12.23	12.40	12.46	12.66	0.03						
<b>MYERS-BRIGGS</b>											
Extraversion	4.73	22.92	17.50	15.33	28.59***	*	*	*			*
Introversion	23.93	4.84	10.50	12.93	26.37***	*	*	*			*
Sensing	28.33	3.23	20.83	8.73	41.58***	*	*		*		*
Intuiting	5.20	22.30	8.08	15.20	31.08***	*	*		*		*
Thinking	23.73	14.07	28.75	10.06	17.13***				*	*	*
Feeling	4.46	10.38	1.33	9.60	11.63***				*	*	*
Judging	1.79	27.69	22.41	14.46	79.12***	*	*	*	*	*	*
Perceiving	13.80	6.08	27.69	1.73	74.44***	*	*	*	*	*	*
<b>KAI</b>											
	101.93	79.15	116.16	58.40	30.17***		*	*			*

\*\*\*  $p < .001$ .\*\*  $p < .01$ .\*  $p < .05$ .

perceive bankers as adaptors. Both the fake advertising executive and the fake librarian score fell within the normal range, as specified by Kirton (1976).

## DISCUSSION

The results of this study indicated as predicted that four personality/preference measures used widely in occupational selection and vocational guidance were highly susceptible to deliberate faking. Of the measures used, the Myers-Briggs seemed most vulnerable to faking, a finding that confirms Furnham (1990), while the VPI seems least susceptible. On the other hand it could be argued that large and numerous differences reflect not so much the measure's fakeability but that the dimensions it

measures are salient for the occupations outlined. Thus some personality trait dimensions might be applicable only to certain jobs and not others.

Some of the results show a fairly predictable pattern—librarians are seen as isolated introverts; advertising executives as sociable, affectionate, artistic, intuiting extraverts; bankers as controlling, conventional, masculine, high status and sensing thinkers; and the subjects themselves as affectionate, scientific, extraverts and innovators (which perhaps indicates that the subjects themselves are faking to some extent). Yet the modal “type” for librarian is ISTP and according to Myers and McCauley (1985) this profile is most associated with such occupations as farmer, the armed forces and the police!

However, the major point of this study was not to explore occupational stereotypes, though it did shed considerable light on this, but to demonstrate that good subjects are quite able not simply to fake, but to fake different, specifically “desirable” profiles for particular purposes. That is, given the task of presenting themselves in a desirable light for a particular job, candidates can, if they so wish, provide a quite specific profile in line with their views or perception of that occupation. Three important caveats should, however, be made: the first concerns the accuracy of the stereotype. By definition, stereotypes have only very generalized accuracy and can frequently be quite wrong; hence it may be that rather than facilitating a candidate’s selection, they actually hinder it. Ideally, one should actually determine the accuracy of the stereotype by getting representatives or incumbents of these actual jobs to complete the questionnaire and see if they respond as the subject supposed they would. Secondly, and perhaps more importantly, there was not a great deal of consensus as to the ideal profile. In fact, the variability scores in the faking groups was as high as, if not higher than, for the control group, showing considerable evidence of disagreement between subjects. It is indeed precisely the fact that there remains so little consensus on the desirable response that many self-report measures remain valid despite faking because the perception of desirability may indeed relate specifically to a subject’s actual score. This assumes that there are individual differences correlated of faking but that faking does not completely threaten the reliability and validity of personality questionnaires. This raises again the big issue of what we mean by “validity.” Clearly, if the test scores bear no relation to the “true” scores, then the test is invalid and the scores cannot be validly used for any purpose. If the test scores are equal to the true scores then the test has perfect validity. If the test scores bear some lawful relationship to the true scores (e.g., test score = true score +  $y\%$ ) then the scores may be validly used for some purposes (e.g., correctly identifying the top  $x\%$  of the population). For some other purposes (e.g., calculating the true population mean) the test scores would be useless and it would be misleading to describe them (or the test) as “valid,” although they would be perfectly reliable.

Third, the argument that subjects can produce the exact profile they were aiming to produce can of course only be confirmed empirically. If indeed we want to know whether each subject produces a specific profile in line with his or her model of what the profile should be, detailed postquestionnaire interviewing needs to take place. Only if they understand the actual dimensions being measured, which items load on



which scales and how the scores are computed could one expect totally accurate profile generation.

Faking studies can however prove useful for at least two reasons. First they can reveal what a subject in general believes to be desirable or normal for a particular group or occupation. The results from the Myers-Briggs, Firo-B and Kirton scales showed some interesting results. For instance, librarians were seen to have a very low judging and high perceiving score, while it was precisely the opposite for advertising agents. Similarly, faking can show an employer what a prospective employee thinks are the most desirable traits for the job, which might or might not conflict with the judgment of employers; or indeed with empirical facts that demonstrate actual traits associated with success and failure in particular jobs. Second, faking studies may provide a useful template of typical faked responses that could be used to actually detect people lying on the questionnaire although one cannot deduce completely, on the basis of the personality scores, whether or not the subject is faking. On the other hand, there is increasing evidence from studies on the consistency and stability of socially desirable responses to suggest that faking may have trait-like qualities that relate to naivety. Furnham (1986a) has argued that the reason why mental health measures are so susceptible to faking, (and correlated with measures of social desirability) is that giving socially desirable responses is, in and of itself an index of mental illness. Thus it is possible that if people fake a "too-good" extremely able, well-adapted profile on application questionnaires they are likely to be mentally unstable or else prone to ingratiation and dissimulation to achieve some end. On the other hand it should be pointed out that subjects who are able to fake skillfully in psychiatric settings are typically better adjusted (Furnham 1988). It is possible that some mild forms of "faking good" are highly appropriate for job applicants in that the total absence of effort at creating distorted impressions may have psychological correlates. This question, however, awaits further empirical investigation.

It should be pointed out that subjects complained of the difficulty of the task, particularly when required to fake "banker." No systematic interviews were performed following the test administration but it does seem as if subjects had difficulty sustaining a consistent fake profile rather than a personal, honest response. It seems therefore that when faking does occur it is not particularly easy to do though the results of this study are very clear-cut: All the tests here were highly susceptible to faking. Furthermore the standard deviation of both faking groups was sufficiently high to indicate that ideas of what constitutes an ideal response are far from clear. It would therefore seem desirable to have a lie scale or measure of social desirability in the test battery to pinpoint those dissimulating for any apparent reason. However the validity of lie scales themselves needs to be established and there is evidence that they are themselves, paradoxically, open to both identification and faking (Furnham and Henderson, 1982).

#### NOTES

Date of acceptance for publication: April 24, 1990. Address for correspondence: Dr. Adrian Furnham, Department of Psychology, University College, London, 26 Bedford Way, London, WC1.

## REFERENCES

- Archer, R., Gordon, R., & Kirchner, F. (1987). MMPI response—set characteristics among adolescents. *Journal of Personality Assessment*, *51*, 506–516.
- Briggs-Myers, L., & McCauley, M. (1985). *A Guide to the development and use of the Myers-Briggs Type indicator*. Palo Alto: Consulting Psychologist Press.
- Burbeck, E., & Furnham, A. (1984). Personality and police selection: Trait differences in successful and non-successful applicants to the Metropolitan Police. *Personality and Individual Differences*, *5*, 257–263.
- Cattell, R. (1969). *The 16 PF, Form D, 1969 Edition*, Windsor: N.F.E.R.
- Furnham, A. (1986A). The Social responsibility of the Type A behaviour pattern. *Psychological Medicine*, *16*, 805–811.
- Furnham, A. (1986B). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, *7*, 385–406.
- Furnham, A. (1986). *Lay Theories*. Pergamon: Oxford.
- Furnham, A. (1990). The fakeability of the 16 PF, Myers-Briggs and FIRO-B personality measures. *Personality and Individual Differences*. In Press.
- Furnham A., & Craig, S. (1987). Fakeability and correlates of the Perception and Preference inventory. *Personality and Individual Differences*, *8*, 459–470.
- Furnham, A., & Henderson, M. (1982). The good, the bad and the mad: response bias in self-report measures. *Personality and Individual Differences*, *3*, 311–320.
- Kirton, M. (1976). Adaptors and Innovators: A description and measure. *Journal of Applied Psychology*, 622–629.
- Kostick, M. (1977). *Kosticks Perception and Preference Inventory*. Brookline, Mass.: Applied Psychology Associates.
- Holland, J. (1973). *Making Vocational Choices: A theory of Careers*. Englewood Cliffs, NJ: Prentice-Hall.
- Holland, J. (1985). *Vocational Preferences Inventory*. Odessa, Florida: PAR Inc.
- Meredith, G. (1968). Stereotypic desirability profiles for the 16 PF questionnaire. *Psychological Reports*, *23*, 1173–1174.
- Myers, I., & McCauley, M. (1985). *Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*. Palo Alto: CPP.
- Robertson, I., & Makin, P. (1986). Management selection in Britain: A survey and critique. *Journal of Occupational Psychology*, *59*, 45–57.
- Velicer, W., & Weiner, B. (1975). Effects of sophistication and faking sets on the Eysenck Personality Inventory. *Psychological Reports*, *37*, 71–73.