

PERSONALITY AND SOCIAL DESIRABILITY IN ORGANIZATIONAL SETTINGS: PRACTICAL IMPLICATIONS FOR WORK AND ORGANIZATIONAL PSYCHOLOGY

Jesús F. Salgado

University of Santiago de Compostela

This article consists of four sections. The first section presents a validity review of personality measures for predicting several organizational criteria, including job performance, training success, leadership emergence, leadership effectiveness, work accidents, job satisfaction, turnover, counterproductive behaviours, absenteeism, and salary. Secondly, the literature on social desirability and distortion of responses to personality questionnaires, one of the main problems of personality measures, is examined. This examination suggests that social desirability has an average effect size of 0.38 standard deviation units on the personality measures, inflating the scores in personnel selection settings. However, social desirability has no effect on the predictive validity of personality measures, and is not a moderator, mediator or suppressor variable. It is also observed that social desirability is not related to job performance. In the third section, the various strategies developed for reducing social desirability are reviewed, and only two are found to be effective: (a) warning applicants that their responses will be checked for social desirability and that distortion will be penalized, and (b) developing norms using job-applicant samples or samples including individuals responding to personality measures in contexts which can produce social desirability (e.g., promotion decisions). The article ends with some conclusions and some suggestions for practitioners in Work and Organizational Psychology.

Este artículo presenta una revisión de la validez de las diversas medidas de personalidad para predecir diversos criterios organizacionales, entre los que se incluyen, el desempeño laboral, el éxito en la formación, la emergencia del liderazgo, la eficacia del liderazgo, los accidentes laborales, la satisfacción laboral, la rotación en el empleo, las conductas contraproducidas, el absentismo y el salario. Seguidamente, se ocupa de examinar uno de los problemas principales a los que se han enfrentado las medidas de personalidad en el trabajo: la deseabilidad social y la distorsión de las respuestas a los cuestionarios. De este examen se desprende que la deseabilidad social tiene un efecto promedio de inflar (o desinflar) las puntuaciones alrededor de 0.38 unidades de desviación en contextos de selección, aunque no afecta a la validez predictiva de las medidas de personalidad, ni es una variable mediadora, moderadora o supresora de la validez de dichas medidas. También se observa que la deseabilidad social no está relacionada con el desempeño en el trabajo. En tercer lugar, se revisan las distintas estrategias utilizadas para reducir y neutralizar la deseabilidad y se observa que sólo dos de ellas son efectivas: (a) informar a los evaluados de que se examinarán sus respuestas en relación con la deseabilidad social y que la distorsión podrá tener consecuencias negativas para los distorsionadores, y (b) crear baremos a partir de muestras de solicitantes o de personas que contestan a las medidas de personalidad en contextos que puedan suscitar la deseabilidad social (p.e., decisiones de promoción). La última parte del artículo contiene las conclusiones y las sugerencias para los profesionales de la Psicología del Trabajo y las Organizaciones.

In the last fifteen years, the assessment of personality has become an issue of great interest for professionals and researchers in the field of Work and Organizational Psychology, given its utility for making decisions in connection with work-related processes. Different meta-analyses carried out in America, Europe, Africa and Asia have reached essentially the same conclusions: personality measures are good predic-

tors of various relevant organizational criteria (see Barrick & Mount, 1991; Barrick, Mount & Judge, 2001; Hogan & Holland, 2003; Hough, 1992; Hertz & Donovan, 2000; Mount & Barrick, 1995; Rothman, Meining & Barrick, 2002; Salgado, 1997; 1998; 2002; 2003; Yoo & Ming, 2002). The fundamental basis of these convergent conclusions is the use of the Five Factor Model (FFM) of personality as a taxonomy for integrating the results of hundreds of local validity studies carried out over more than 60 years. In accordance with this model, five broad dimensions of personality have been found to be replicable across different samples, in different cultures, for dif-

Correspondence: Jesús F. Salgado. Departamento de Psicología Social, Básica y Metodología, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain.
E-mail: psjesal@usc.es

TABLE 1
SUMMARY OF META-ANALYTICAL RESULTS ON THE
RELATIONSHIP BETWEEN PERSONALITY MEASURES AND
VARIOUS ORGANIZATIONAL CRITERIA AND VARIABLES
 (Original source: Salgado & De Fruyt, 2005)

Dimension	K	N	Validity
Job Performance ^a			
Conscientiousness	133	33,668	.33
Emotional Stability	108	19,880	.21
Extraversion	111	21,916	.10
Openness to Experience	82	13,895	.09
Agreeableness	110	21,911	.19
Training Proficiency ^b			
Conscientiousness	20	3,909	.31
Emotional Stability	25	3,753	.09
Extraversion	21	3,484	.28
Openness to Experience	18	3,177	.33
Agreeableness	24	4,100	.14
Leadership Emergence ^c			
Conscientiousness	17	n/a	.33
Emotional Stability	30	n/a	.24
Extraversion	37	n/a	.33
Openness to Experience	20	n/a	.24
Agreeableness	23	n/a	.05
Effective Leadership ^c			
Conscientiousness	18	n/a	.16
Emotional Stability	18	n/a	.22
Extraversion	23	n/a	.24
Openness to Experience	17	n/a	.24
Agreeableness	19	n/a	.16
Job Satisfaction ^d			
Conscientiousness	79	21,719	.26
Emotional Stability	92	24,527	.29
Extraversion	75	20,184	.25
Openness to Experience	50	15,196	.02
Agreeableness	38	11,856	.17
Counterproductive behaviours (Validity reversed) ^e			
Conscientiousness	13	6,276	.26
Emotional Stability	15	3,107	.06
Extraversion	12	2,383	-.01
Openness to Experience	8	1,421	-.14
Agreeableness	9	1,299	.20
Turnover (Validity reversed) ^e			
Conscientiousness	5	748	.31
Emotional Stability	4	554	.35
Extraversion	4	554	.20
Openness to Experience	4	554	.14
Agreeableness	4	554	.22
Accidents at work (Validity reversed) ^f			
Conscientiousness	9	1125	.30
Emotional Stability	13	1198	.28
Extraversion	12	1524	-.09
Openness to Experience	7	570	-.50
Agreeableness	7	420	.61

Note: K= number of studies; N= Total sample size; n/a= not available; ^a Salgado (2004); ^b Barrick, Mount & Judge (2001); ^c Judge, Bono, Ilies & Gerhardt (2002); ^d Judge, Heller & Mount (2002); ^e Salgado (2002); ^f Clarke & Robertson (2005).

ferent languages and with different assessment techniques. The names of these factors vary among the different researchers in the field of the psychology of personality in the work context, but the labels most widely used are those suggested by Costa and McCrae (1992). For them, the five personality dimensions or factors would be: emotional stability (versus neuroticism), extraversion (versus introversion), openness to experience (versus closure to experience), agreeableness (versus antagonism) and conscientiousness (versus irresponsibility). The names of the emotional stability and extraversion factors are those on which there is greatest consensus. The other three, perhaps because they are more recent, are denoted by a variety of terms. For example, the openness to experience dimension has also been called culture or intellectuality; the agreeableness dimension has been called friendliness, and the conscientiousness dimension has been labelled dependability, prudence, responsibility or need for achievement. The results of the meta-analyses mentioned above have shown that two personality factors, emotional stability and conscientiousness, are valid predictors of job performance in all occupations. It has also been shown that these two factors, together with extraversion, are valid predictors of training proficiency, and that the conscientiousness and agreeableness factors predict counterproductive behaviour.

In addition to the above, research on personality in organizational settings has demonstrated that personality measures focusing on occupational criteria (*Criterion-focused Occupational Personality Scales, COPS*), such as integrity tests, client orientation scales, management potential scales, stress tolerance scales or commercial potential scales are excellent predictors of diverse organizational criteria, including job performance, training proficiency and counterproductive behaviours (see Ones & Viswesvaran, 2001a and b, for a summary). Research has also shown that these measures (COPS) are a combination of three of the basic personality factors: emotional stability, agreeableness and conscientiousness. Tables 1, 2 and 3 provide a summary of various personality dimensions and composites for the prediction of different organizational criteria.

In the light of the results of the meta-analytical studies mentioned, it is clear why personality measures have been so extensively used in organizational decisions in the last decade, and why they have aroused the interest of professionals. Although they have primarily been used

for purposes of personnel selection, they have also been employed in training processes and personnel development (e.g., *coaching*), and for establishing competence profiles.

Nevertheless, and despite this recent success of personality measures for predicting job performance, their use in organizational contexts is not without its problems and difficulties. Among these problems, distortion and response bias are two of those that have received most interest, and about which there has been most concern; consequently, there has been an abundance of studies dealing with these issues in recent years – though it is in fact sixty years since Mehl and Hathaway (1946) and Ellis (1946) demonstrated that people, when taught to do so, can distort responses to personality questionnaires.

DISTORTION OF RESPONSES TO ITEMS OF PERSONALITY MEASURES

Distortion and response bias to personality measures items, especially in organizational contexts, where deci-

sions based on responses to personality questionnaires have important implications for the respondents (e.g., being hired or not), have received considerable attention from research over the last 50 years or more. Response distortion can be either positive (e.g., trying to make a good impression) or negative (e.g., trying to suggest a psychological problem or disorder). In the context of organizations, although the latter type of bias exists in certain circumstances, it is the former type, positive distortion, that has given the most cause for concern, given that personality measures have been used primarily for purposes of personal selection, so that ‘trying to give

TABLE 2
SUMMARY OF META-ANALYTICAL RESULTS ON THE RELATIONSHIP BETWEEN COPS AND VARIOUS ORGANIZATIONAL CRITERIA AND VARIABLES
(Original source: Salgado & De Fruyt, 2005)

Dimension	K	N	Validity
Job Performance			
Integrity (personality tests) ^a	102	27,081	.37
Drugs and Alcohol Scales ^b	7	1,436	.19
Stress Tolerance Scales ^b	13	1,010	.42
Client Orientation Scales ^b	33	6,944	.39
Violence Scales ^c	14	4,003	.41
Training Proficiency			
Integrity (personality tests) ^d		2,364	.38
Counterproductive behaviours (Validity inverted) ^e			
Integrity (personality tests) ^a	138	158,065	.32
Stress Tolerance Scales ^b	5	594	.42
Client Orientation Scales ^b	5	740	.42
Violence Scales ^c	4	533	.46
Absenteeism			
Integrity (personality tests) ^e	16	5,435	.36

Note: K= number of studies; N= Total sample size; ^a=Ones, Viswesvaran & Schmidt (1993); ^b= Ones & Viswesvaran (2001a); ^c= Ones & Viswesvaran (2001b); ^d= Ones & Viswesvaran (1998a); ^e= Ones, Viswesvaran & Schmidt (2003).

TABLE 3
SUMMARY OF THE META-ANALYTICAL RESULTS BETWEEN SEVERAL PERSONALITY VARIABLES (AND MODELS) AND SEVERAL ORGANIZATIONAL CRITERIA AND VARIABLES.
(Original source: Salgado & De Fruyt, 2005)

Dimension	K	N	Validity
Job Performance			
Conscientiousness-NFFM ^a	36	5,874	.18
Emotional Stability-NFFM ^a	25	4,541	.05
Extraversion-NFFM ^a	26	4,338	.08
Openness to Experience-NFFM ^a	29	4,364	.08
Agreeableness-NFFM ^a	31	4,573	.13
Generalized Self-Efficacy ^b	11	1,506	.43
Locus of Control ^c	35	4,310	.22
Self-Esteem ^c	40	5,145	.26
Training Proficiency			
Generalized Self-Efficacy ^b	4	422	.29
Job Performance			
Emotional Intelligence ^d	19	2,652	.24
Job Satisfaction			
Positive Affectivity ^e	15	3,326	.49
Negative Affectivity ^e	27	6,233	-.33
Generalized Self-Efficacy ^c	8	1,411	.29
Locus of Control ^c	80	18,491	.32
Self-Esteem ^c	56	20,819	.26
Salary			
Generalized Self-Efficacy ^b	5	468	.28
Absenteeism			
Generalized Self-Efficacy ^b	4	718	.21

Note: K= number of studies; N= Total sample size; NFFM= measures from questionnaires not based on the Five Factor Model; ^a=Salgado (2003); ^b=Salgado & Moscoso (2000); ^c=Judge & Bono (2001) ^d=Van Rooy & Viswesvaran (2004); ^e=Connolly & Viswesvaran (2000).

a good impression' is frequently an option as a response to tests and questionnaires. In view of the fact that response distortion in a favourable direction can favour candidates' possibilities of being hired, Seisdedos (1988), in one of the few studies carried out in Spain on this phenomenon in organizations, has called this bias "Intelligent adaptation". However, it is typically labelled as 'Social desirability', 'Sincerity' or 'Motivational Distortion' – though other authors refer to it as 'Faking', 'Infrequent Virtues' or 'Response Distortion'. Of all these terms, 'Social desirability' is that most commonly used, and covers all the others.

Social desirability (or response distortion) has been defined as "the tendency to endorse items in response to social or normative pressures instead of providing veridical self-reports" (Ellingson, Smith & Sackett, 2001, p.122). Although it has often been seen as a unidimensional construct, and many measures provide only a global measure of it, according to the most recent research, mainly that of Paulhus (1984, 2002), desirability has two dimensions, which have been called *impression management* and *self-deception*. Impression management indicates a tendency to intentionally adapt one's public image with the aim of being favourably viewed by others. Self-deception, on the other hand, refers to the unintentional tendency to describe oneself in a favourable manner, and is expressed through self-descriptions that are positively biased but in which one honestly believes. In this sense, impression management is a voluntary manipulation of one's own image so that others perceive us in a positive way, whilst self-deception is not a deliberate manipulation, though it may lead to distortions in others' perceptions of us. Taking into account this distinction between the two dimensions of social desirability, in the context of the Psychology of Work and Organizations, the effects of impression management on personality measures scores appear to be the more relevant.

The distinction between impression management and self-deception is relevant in the light of the suggestion by some researchers in the psychology of personality that there are individual differences in social desirability (Block 1965; McCrae & Costa, 1983). This means that social desirability may be not simply a tendency to adapt to situations, but rather a stable personality characteristic that would indicate more substantive and significant differences. For example, McCrae and Costa, (1983), Ones, Viswesvaran and Reiss (1996), and Salgado, Igle-

sias and Remeseiro (1996) have found that social desirability correlates with emotional stability and with conscientiousness. As McCrae and Costa (1983) point out, this would mean that a person who genuinely scores highly in conscientiousness, and who is emotionally stable and cooperative (scoring highly in agreeableness) will also score highly in social desirability. However – and this is the paradox –, this person may be honest and reliable, but would be "guilty" of distortion or lying in personality questionnaires.

The effects of social desirability on personality measures and on other assessment instruments (e.g., interviews) are widely documented in the literature, and are particularly well known to those responsible for personnel selection in organizations. The principal effect is that social desirability tends to inflate (increase) scores in the dimensions that are (or that candidates believe to be) positively related to job performance and to deflate (reduce) scores in those dimensions that are (or that candidates believe to be) negatively related to job performance. Recent research has reviewed meta-analytically, and through studies with large samples, the effects of social desirability on responses to personality measures. In general, such empirical research has consisted in the use of three types of design: (a) comparisons between groups in laboratory situations, with participants instructed for distorting their responses; (b) within-subject comparisons in laboratory situations, with participants instructed to distort, and (c) comparisons in real selection situations, examining the differences between those who show social desirability and those who do not.

Ones and colleagues (Ones & Viswesvaran 1998a and b; Ones, Viswesvaran & Reiss, 1996; Viswesvaran, Ones & Hough, 2001), Hough and colleagues (1998; Hough, Eaton, Dunnette, Kamp & McCloy, 1990; Hough & Paullin, 1994) and Christiansen and colleagues (Christiansen, Goffin, Johnston & Rothstein, 1994; Goffin & Christiansen, 2003) are those that have done most research on the effects of social desirability in organizational settings, though other researchers have made relevant contributions. The main results are as follows. In the studies with comparisons between groups in laboratory situations with induced positive distortion it was found that, on questionnaires measuring personality dimensions (e.g., the *Big Five*), "fakers" score, on average, 0.6 standard deviation units more than "non-fakers". In terms of T scores (mean=50; SD=10), this means that the fakers group would score an average of 56, as against a



score of 50 for the non-fakers group. When studies make within-subject comparisons in laboratory situations with induced distortion and honest response, the difference in the scores on personality dimensions measures between the two conditions for the same participants is 0.72 standard deviation units, which in terms of T scores means that the distortion increases a person's score from 50 to 57.2 points. When type of design is not taken into account, the difference is 0.5 standard deviation units. Thus, these results indicate that social desirability has effects on distorters' scores of slightly over half a standard deviation unit. They also show that type of research design has important effects on the size of the distortion. This same type of study was carried out with composite personality measures (COPS), such as integrity tests, with quite similar results, since the difference between people instructed to respond in a positively distorted manner and those instructed to respond honestly was 0.50 standard deviation units.

The results described above can be considered as indicative of the effects of social desirability in situations of maximum performance (when the aim is to find the greatest possible difference between the conditions of social desirability and honesty). However, studies on the effects of social desirability in "real" conditions, that is, which compare the responses of fakers and non-fakers in genuine selection situations, would reflect the typical performance situation (where the aim is to find the usual difference between the two mentioned conditions in a typical work context). In this regard, Hough (1998) carried out three particularly relevant studies. In the first of these she compared a sample of 963 telecommunications employees with a sample of 14,442 job applicants in the same industry, finding in a measure of responsibility (one of the sub-dimensions of the conscientiousness factor) an average distortion of 0.45 standard deviation units on the side of the applicants. In a second study, in relation to positions in the local police, with a sample of 508 employees and 24,433 applicants, using a measure of conscientiousness, Hough found the average distortion to be 0.33 standard deviation units on the side of the applicants. A third study, with a sample of 270 national guards and 681 applicants, found an average distortion of 0.13 on a personality scale composed of the experience-seeking and self-esteem dimensions. Other studies carried out by Hough, Eaton, Dunnette, Kamp and McCloy (1990) produced similar results. Therefore, on the whole, the available data indicate that in the third type of

research design (comparison of employees and applicants in real situations), the average effects of social desirability, while they exist, are much smaller than those found in laboratory settings. In other words, in situations of typical performance (e.g., personnel selection), the effects of social desirability are less than those found in situations of maximum performance (e.g., experimental manipulation).

In addition to the effect of inflating scores on personality measures, a second possible effect often mentioned, and which is probably that which causes most concern among Work and Organizational Psychology professionals using personality measures for personnel management (hiring decisions, training, promotion, etc.), is that associated with the impact of social desirability on the validity of such measures. Some researchers, and many professionals, believe that social desirability has a negative effect on the validity of the measures, reducing it, and thus making them invalid. In other words, the validity of personality measures for predicting job performance and other organizational criteria would be substantially reduced, or even cancelled out, by the effects of desirability (Goffin & Christiansen, 2003; Mueller-Hanson, Heggstad & Thornton, 2003; Rosse, Stecher, Miller & Levin, 1998). This belief has led to serious doubts over the use of personality measures in organizational settings, and even to some professionals openly rejecting their utility for decision-making.

In relation to this, meta-analytical and individual studies with large samples ($N > 1200$) have recently been carried out to explore the effects of social desirability on the validity of personality measures. Thus, for example, with regard to construct validity, Ones, Viswesvaran and Reiss (1996) showed that social desirability has scarcely any effect on the convergent and discriminant validity of personality measures, the average increase in correlations being of the order of .015, that is, irrelevant in practice. As far as criterion (predictive) validity is concerned, Ones and Viswesvaran (1998b) have examined the effects of social desirability (whether it be considered as a moderator, mediator or suppressor variable). The results of their meta-analyses indicate that social desirability has no impact on the validity of personality measures, whether these refer to the basic dimensions (e.g., the *Big Five*) or personality composites (e.g., integrity), with validity remaining essentially the same once the effects of social desirability have been accounted for. The studies by Hough (1998; Hough et al., 1990) obtained



results that were basically similar. Such work has demonstrated the erroneous nature of the beliefs of some researchers and many professionals about the negative effects of social desirability on the validity of personality measures.

HOW IS SOCIAL DESIRABILITY DETECTED (MEASURED)?

Given the interest generated by social desirability and its effects on personality measures scores, several researchers have considered the question of how to detect it; consequently, over the years, a series of instruments have been designed, though almost all of them take as their source the *Minnesota Multiphasic Personality Inventory* (MMPI). The developers of the MMPI, Hathaway and McKinley, actually designed two scales for detecting possible distortion in responses to the questionnaire. One is the so-called K scale, for detecting negative bias, or the tendency to present a poorer image of oneself; the other is the *Lie* (L) scale, sometimes described as the Sincerity scale, and which is aimed at detecting positive bias, or the tendency to project a good image of oneself. Both scales were developed as criterion-focused tests, as the rest of the MMPI scales had been. Following the line of the MMPI, Harrison Gough, creator of the *California Personality Inventory* (CPI; 1987) and a student of Hathaway, also developed a scale for measuring social desirability in his questionnaire, calling it the *Good Impression Scale*. Given that the CPI was designed to assess normal (adjusted) personality, in contrast to the MMPI, which was intended for the assessment of personality disorders, Gough was most interested in the tendency to distort positively. Eysenck, in his first personality questionnaire, the *Maudsley Personality Inventory* (MPI; predecessor of the *Eysenck Personality Inventory*, EPI), also included a scale for assessing social desirability. This scale was also derived from the MMPI's L scale. Likewise, many later researchers have created scales for measuring social desirability, and a good deal of personality questionnaires include among their items some scale or other measure in relation to such distortion. In addition to those mentioned above, other popular instruments include Edwards' social desirability scale (1957), Crowne-Marlowe's social desirability scale (1964), Eysenck's sincerity scale (Eysenck & Eysenck, 1964; originally a lying scale), the positive motivational distortion scale of the 16PF (Cattell, Eber & Tatsuoka, 1970) or the social desirability scale of the Occupational Personality Questionnaire (SHL, 1999). All of these scales were designed on

the basis of social desirability being a unidimensional concept. However, since the research by Paulhus referred to above, it has been accepted that social desirability comprises two dimensions, and the *Balanced Inventory of Desirable Responding* (BIDR, Paulhus 1984, 2002) is the most popular of the modern inventories. This inventory consists of 40 items, with two 20-item subscales, which assess 'impression management' and 'self-deception'.

Nevertheless, it should be pointed out that many recent personality questionnaires, especially those based on the Five Factor model, do not use a scale for detecting social desirability. This is the case, for example, of the NEO-PI-R (Costa & McCrae, 1992), the *Hogan Personality Inventory* (HPI, Hogan & Hogan, 1995), the IP/5F (Salgado, 1996) or the *Jackson Personality Inventory* (JPI, Jackson, 1994).

STRATEGIES FOR REDUCING THE EFFECTS OF SOCIAL DESIRABILITY

Having examined the effects of social desirability in organizational contexts, and seeing that they involve the inflation or deflation of scores on measures of some dimensions and facets of personality, the following question concerns how these effects can be reduced or cancelled out.

In this regard, over the years a variety of strategies have been proposed. Among them are the following: (a) use of scales with forced-choice items, the items having been paired according to their similarity in social desirability (this mode is also called 'ipsative measure' or 'ipsative score'); (b) use of scales for detecting social desirability and discarding of respondents who score moderately high (e.g., 2 standard deviation units over the mean) on these scales; (c) use of detection scales and subjective "adjustment" of the personality-measure scores of respondents scoring moderately high on the social desirability scales; (d) use of detection scales and mechanical "correction" of the personality-measure scores of respondents who distort moderately highly, and use of a mathematical formula for this purpose; (e) warning respondents about the existence of detection methods in the assessment, and instructing them about the possible consequences of distortion; (f) developing specific norms for samples of applicants, rather than using the norms of normative samples from the general population. Table 4 provides a summary of these strategies.

Currently, there is very little use of scales with forced-choice format for assessing personality in organizational



contexts, especially for purposes of personnel selection. Nevertheless, there are some questionnaires that use this format, with the intention of reducing distortion in responses. Examples of personality questionnaires that employ this approach are the *Occupational Personality Questionnaire 3.2* (OPQ 3.2; SHL, 1999), the *Thomas Personality Inventory*, also known as the DISC (Thomas International), or the *Description en Cinq Dimensions* (D5D, Rolland & Mogenet, 2001). The basic assumption of those who design this type of questionnaire is that if the items are grouped, for example, in fours, with similar social desirability, and respondents have to indicate which of the items best defines them and which is the least appropriate (discarding the other two options), then the final response will better reflect their personality characteristics and will eliminate the effects of social desirability. Hicks (1970, p.181) suggested that to justify the use of forced-choice (ipsative) measures, three conditions were required: (1) that there is a marked bias in responses to personality questionnaires, (2) that this bias reduces the validity, and (3) that the forced-choice format reduces the bias and increases the validity to a greater extent than other, non-ipsative controls of bias. Hicks concluded that no case had occurred in which these three conditions were jointly met. Twenty-five years later, Bartram (1996) considered that Hicks' conclusion remained true. In this regard, the following should be noted: (a) the most recent and exhaustive research, mentioned here in previous sections, has demonstrated that there is a bias, and that it can be important in real assessment situations in organizational contexts (e.g., personnel selection); and (b) research has also shown that social desirability does not reduce validity. As regards the third condition, referring to the fact that forced-choice (ipsative) formats reduce bias and increase validity, various studies have dealing with this issue have appeared in recent years. For example, Christiansen, Burns and Montgomery (2005) carried out a series of studies showing that (1) forced-choice scales are as susceptible to distortion as traditional normative scales, and (2) respondents with higher scores in measures of general mental ability are more successful in improving their scores on forced-choice scales (in the sense of more closely fitting the desired profile), suggesting that such scales are manipulable according to respondents' general mental ability. Baron (1996), a staunch defender of this format for personality measures, acknowledges that a small group of candidates powerfully distort scores on these

questionnaires. Consequently, Hicks' third condition, in the light of recent data, is not fulfilled in practice. Thus, and bearing in mind that this response format has many and serious limitations of a psychometric nature, with regard to its reliability, its factorization, its validity and comparisons between individuals (see, for example, the reviews by Baron, 1996; Bartram, 1996; Closs, 1996; Cornwell & Dunlap, 1994; Hicks, 1970; and Meade, 2004), the decision on this strategy is that it should be discarded, and not used for purposes of assessment in organizational contexts involving the comparison of persons among one another (e.g., selection).

The second strategy, discarding those respondents who score highly on social desirability scales, introduces serious complications of a theoretical, practical and possibly even legal nature. From the theoretical point of view, it is possible, as McCrae and Costa (1983) point out, that persons who are totally honest in their responses, but with certain personality characteristics that fit the typical profile of the distorter, would be excluded without further consideration. This would be a clear case of 'false positives', which would be rejected as a result of this strategy. Furthermore, and also from the theoretical perspective, it would be necessary to demonstrate that those scoring highest on motivational distortion scales subsequently present a job performance inferior to that of non-distorters. And not only has this not been demonstrated, but it has also been shown that there is no relationship between social desirability and job performance (Ones & Viswesvaran, 1998b). Nor has it been demonstrated that respondents who distort, and consequently present higher scores in some personality dimensions (e.g., conscientiousness) related to job performance, later show (after being hired) poorer job performance than those who obtain similar scores without distorting their responses. From the practical point of view, the exclusion of a number of persons from the set of respondents may be counter-indicated if the selection ratio is high (e.g., it approaches 1) – that is, if the numbers of vacancies and of candidates are similar. Finally, from the legal perspective, the exclusion of candidates on the basis of their distorted responses cannot be defended (in court, for instance) when it is known, as is now the case, that there is no negative relationship between social desirability and job performance. Therefore, having distorted cannot legally be a reason for exclusion, and any candidate lodging an appeal against a decision based on this point would have a very good chance of a favourable verdict.



Consequently, this strategy should not be used by professionals from the field of Work and Organizational Psychology in certain tasks carried out in labour-related contexts.

The third and fourth strategies consist in using detection scales and subjectively “adjusting” the scores on the personality measures of respondents who score moderately highly on the social desirability scales, or mechanically “correcting” the personality-measure scores of moderately high distorters, using a mathematical formula. As it can be appreciated, the two strategies are quite alike, and involve similar problems. Correction of scores, increasing or reducing them according to the degree of distortion detected on the social desirability scale, has

been a widely used strategy among psychologists in general and those from the field of work and organizations in particular. This is due to the fact that two of the most popular personality questionnaires – MMPI and the 16PF – include systems for ‘correcting’ the scores. In the case of these two questionnaires, the strategy is the fourth one, that is, mechanical correction (based on a regression equation derived from motivational distortion), but it has led to many professionals taking a subjective approach to correction, based on their experience and the evaluator’s ‘clinical eye’, and to their adjusting the scores accordingly. A problem common to the two methods, and usually overlooked by professionals, is that correction of the scores modifies the construct validity of the question-

TABLE 4
STRATEGIES FOR THE REDUCTION OF SOCIAL DESIRABILITY AND SUGGESTIONS FOR PROFESSIONALS

Strategy	Description	Limitations	Effectiveness	Recommendation
Forced-Choice Scales	Choose between items with similar degree of social desirability	Theoretical, methodological and practical	Limited	Not recommended
Use Social Desirability scales and discard candidates scoring highly on them	Include a measure of distortion and exclude those who score above a cut-off point	Theoretical, practical and legal	Ineffective	Not recommended
Use Social Desirability scales and adjust distorters’ scores (subjective strategy)	Scores of candidates considered to be fakers are adjusted “subjectively”, based on the assessor’s experience	Theoretical and practical	Ineffective	Not recommended
Use Social Desirability scales and adjust distorters’ scores (objective strategy)	Scores of candidates considered to be fakers are adjusted “subjectively”, based, for example, on a regression equation	Theoretical and practical	Ineffective	Not recommended
Warn candidates	Candidates are warned of the possibility of being eliminated or penalized if they distort their responses	None	Effective	Recommended
Develop norms based on samples of job applicants	Calibrate candidates’ scores after creating a norm developed with samples of job applicants, rather than a sample of the general population	None	Effective	Recommended



naires (see Ones & Viswesvaran, 1998a and b), which means that the measure and its reliability are modified, without improvement to its criterion (predictive) validity, which may indeed be adversely affected (see, Hough 1998). Thus, scores derived from the correction and adjustment of raw data may fail to correspond to the respondent's actual personality characteristics. Moreover, the third strategy is practically unfeasible when dealing with a large number of candidates, since it requires the examination of each particular profile in order to carry out the correction. Consequently, these two strategies are also at odds with sound professional practice based on current empirical knowledge. It is not surprising, therefore, that in the latest version of the 16PF the correction of scores is no longer used.

The fifth strategy referred to above consists in warning respondents to the personality questionnaires that these incorporate methods for detecting social desirability and other possible biases, and asking them to be as sincere and honest as possible. At the same time, they are warned that those respondents detected as fakers may be disregarded as candidates for further consideration, or penalized in some other way appropriate to the case. For example, Hough (1998) used this strategy in several selection processes, warning applicants that those who distorted their scores on the personality questionnaire would be detected, and that those who were identified as having provided exaggeratedly favourable self-descriptions would be disqualified from the selection process for six months, after which time they could reapply for assessment. Although Hough did not assess the effectiveness of this fifth strategy (being more interested in comparing the difference in bias between employees and applicants), from the data she provides (Hough, 1998) and those of Ones, Viswesvaran and Reiss (1996), it is possible to estimate its effectiveness. Hough's three studies (1998) have an accumulated total sample of 40,297 persons, and average distortion weighted by number of applicants in relation to number of employees is 0.37 standard deviation units. Bearing in mind that the present case involves comparing fakers and non-fakers, we can use as an estimation of this comparison that obtained by Ones et al. (1996) for comparisons between groups in situations of maximum performance, which was 0.60 standard deviation units. Thus, deducting 0.37 standard deviation units corresponding to fakers in selection situations from the figure of 0.60 (the maximum distortion that can be obtained), the resulting value is 0.23 standard

deviation units less, which is directly attributable to the strategy of warning the candidates. The results of these two studies, then, indicate that this strategy is effective in the reduction of social desirability, bringing it down by more than 38%. Moreover, given that the penalization suggested by Hough merely postpones the decision on hiring distorter candidates, this strategy is legally defensible. More recently, McFarland (2003) examined, in a laboratory setting, the effects of this strategy for reducing distortion, and assessed its effects on applicants' reactions in relation to the perceived organizational justice. McFarland's results indicate that the strategy was highly effective in reducing social desirability (0.45 standard deviation units on average), and that, moreover, it had no negative effects on applicants' perceptions (in one case these were even positive) and reduced multicollinearity between the personality variables. Consequently, this strategy is both valid (it reduces distortion) and economical, as there is no need to actually include a measure of social desirability – it is sufficient to inform candidates that there is one and that it can lead to their being penalized.

The final strategy mentioned consists in developing specific norms for contexts of personnel selection. In other words, instead of using the norms corresponding to the general population or to particular groups developed on the basis of scores obtained in situations where respondents have no interest in distorting, this strategy involves developing norms from scores obtained in situations where people have a direct interest in distorting (e.g., in personnel selection situations). It is obvious that the norms drawn up in this way will already include a part of the normative score corresponding to distortion, and which is common to all respondents. Thus, the social desirability will have already been partially neutralized. That is, a framework will have been developed and applied that is common to all respondents, rather than using a framework that is only common to those who never distort, such as one based on norms created with a normative population that responded to the questionnaire in a situation of null social desirability (e.g., with samples of students not induced to distort).

CONCLUSIONS AND SUGGESTIONS FOR PRACTICE

The possibility of responses to personality questionnaires being distorted, either positively or negatively, is a real one, and the phenomenon, which has been known of for more than sixty years, continues to be of great concern to



professionals in the field of Work and Organizational Psychology with responsibilities in the use of personality measures in their daily work (e.g., personnel selection). Such concern has led to different approaches to the use of personality measures among these professionals, the three most common being: (a) given the possibility of distortion in responses, to reject the use of personality measures, considering that such distortion invalidates their predictive capacity, so that appropriate decisions cannot be made on the basis of these instruments. The answer in this case has been to remove them from the toolbox of the organizational psychologist; (b) despite acknowledging the potential for distortion, to consider that personality measures continue to be valid and useful for professional work, and to seek formulas for overcoming or neutralizing this deficiency. This is the case of psychologists who have used corrective measures for the scores; (c) to consider that social desirability is not a significant problem, and that it would be much worse to go without the use of personality measures in one's professional practice.

Studies carried out in the last 20 years, and especially in the last ten years, have permitted researchers to reach sound conclusions on the effects of social desirability and possible ways of offsetting it. These conclusions can be summarized as follows:

- 1) Personality measures – those based on the Big Five factor model, criterion-focused occupational personality questionnaires, and instruments based on alternative models – are excellent predictors of job performance, training proficiency, counterproductive behaviours, leadership, job satisfaction, organizational commitment, knowledge acquisition, and many other relevant criteria for organizations. In some cases, personality measures are the best predictors of such criteria (e.g., counterproductive behaviours or job satisfaction).
- 2) Social desirability affects all personality assessment methods based on questionnaires, including those designed to be free of its effects, such as forced-choice questionnaires or ipsative measures. That is, no self-report personality measure is immune from the effects of social desirability.
- 3) Social desirability, as a relevant form of distortion, affects only a small percentage of those assessed in organizational processes.
- 4) The effect of social desirability varies depending on the way such desirability is triggered. In typical se-

lection situations, desirability has an average effect of 0.38 standard deviation units. In terms of T scores, this means an increase or decrease in scores of 3.8 points.

- 5) Social desirability comprises two factors, impression management and self-deception. In organizational contexts the former is the more relevant, and that which inflates (or deflates) scores on personality questionnaires.
- 6) Social desirability is not related to job performance, and therefore does not affect the validity of personality measures. In other words, if the effects of social desirability are eliminated from personality measures, these do not have higher validity for predicting job performance. Therefore, the belief mentioned above, referring to the notion that social desirability invalidates the predictive capacity of personality measures, is simply erroneous.
- 7) Over the years, diverse strategies have been developed for neutralizing or reducing the effects of social desirability, and the majority have turned out to be ineffective or ill-advised:
 - a) The strategy of forced-choice items does not prevent them from being distorted, does not improve the predictive validity of the measures, negatively affects their reliability, involves serious psychometric problems, and does not permit comparisons between respondents, so that it cannot be used appropriately in those situations that require the comparison of candidates (e.g., selection, training). Therefore, it should not be used.
 - b) The strategy of discarding candidates who score highly on the social desirability scales has serious theoretical, practical and legal complications. Therefore, it should not be used.
 - c) The strategy of "correcting" scores in a subjective manner has theoretical and practical complications. It affects the construct and predictive validity of the instruments, and its use is unfeasible with large numbers of candidates. Therefore, it should not be used.
 - d) The strategy of "correcting" scores mechanically (e.g., by means of a regression equation) involves the same theoretical complications as the previous strategy. Therefore, it should not be used.
 - e) The strategy of warning candidates about the existence of detectors of distortion and alerting

them to the potential consequences of such distortion considerably reduces distortion, and is an economical means of controlling social desirability. It should be used in all processes and assessments in which social desirability may be present.

f) The strategy of creating norms based on samples of job applicants neutralizes the effects of social desirability. Therefore, it should be used in all selection processes. And in these cases professionals should use instruments that provide such norms.

8) The combination of strategies of warning the candidates and using norms developed with samples of job applicants produces optimum results for the reduction of social desirability. This is the best option for professionals.

The above conclusions are based on the evidence currently available, chiefly obtained from meta-analyses and studies with large samples ($N > 1200$). In the light of this evidence, professionals in the field of Work and Organizational Psychology can confidently use personality measures for making decisions, with the limitations inherent to any psychological measure (i.e., reliability, validity and utility), without considering social desirability as a problem that invalidates them. Indeed, it seems that all the fuss over social desirability has resulted from the artificial creation of a problem with scarce relevance for the profession.

REFERENCES

- Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, *69*, 49-56.
- Barrick, M.R. & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1-26.
- Barrick, M.R., Mount, M.K. & Judge, T. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*, 9-30.
- Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology*, *69*, 25-39.
- Block, J. (1965). The challenger of response sets: *Unconfounding meaning, acquiescence, and social desirability in the MMPI*. New York: Appleton-Century Crofts.
- Cattell, R.B., Eber, D. & Tatsuoka, M. (1970). *Handbook of 16PF*. Champaign, IL: IPAT.
- Christiansen, N.D., Burns, G.N. & Montgomery, G.E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*, 267-307.
- Christiansen, N.D., Goffin, R.D., Johnston, N.G. & Rothstein, M.G. (1994). Correcting the Sixteen Personality Factors test for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology*, *47*, 847-860.
- Clarke, S. & Robertson, I.T. (2005). A meta-analytic review of the Big Five personality factors and accidents involvement in occupational and non-occupational settings. *Journal of Occupational and Organizational Psychology*, *78*, 355-366.
- Closs, S.J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational and Organizational Psychology*, *69*, 41-47.
- Cornwell, J.M. & Dunlap, W.P. (1994). On the questionable soundness of factoring ipsative data: A response to Saville and Wilson (1991). *Journal of Occupational and Organizational Psychology*, *67*, 89-100.
- Costa, P. T. & McCrae, R. R. (1992). *Professional Manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor-Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.
- Crowne, D.P. & Marlowe, D. (1964). *The approval motive*. New York: Wiley.
- Connolly, J.J. & Viswesvaran, C. (2000). The role of affectivity in job satisfaction: A meta-analysis. *Personality and Individual Differences*, *29*, 265-281
- Edwards, A.L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Ellington, J., Sackett, P.R. & Hough, L.M. (1999). Social desirability corrections in personality measurement: issues of applicant comparison and construct validity. *Journal of Applied Psychology*, *84*, 155-166.
- Ellington, J.E., Smith, D.B., & Sackett, P.R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology*, *86*, 122-133.
- Ellis, A. (1946). The validity of personality questionnaires. *Psychological Bulletin*, *43*, 385-440.

- Eysenck, H.J. & Eysenck, S.B. (1964). *The Eysenck Personality Inventory manual*. London: Routledge.
- Goffin, R.D. & Christiansen, N.D. (2003). Correcting personality tests for faking: A review of popular personality tests and initial survey of researchers. *International Journal of Selection and Assessment*, 11, 340-344.
- Gough, H.G. (1987). *Manual: The California Personality Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Hicks, L.E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.
- Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, 88, 100-112.
- Hough, L.M. (1992). The "Big Five" personality variable construct confusion: Description versus prediction. *Human Performance*, 5, 139-155.
- Hough, L.M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11, 209-244.
- Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D. & McKloy, R.A. (1990). Criterion-related validities of personality constructs and the effects of response distortion on those validities. *Journal of Applied Psychology*, 75, 581-595.
- Hough, L.M. & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G.S. Stokes, M.D. Munford & W.A. Owens (Eds.), *The biodata handbook: Theory, research and application*. Palo Alto, CA: Consulting Psychologists Press.
- Hurtz, G.M. & Donovan, J.J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology*, 85, 869-879.
- Jackson, D.N. (1994). *Jackson Personality Inventory. - Revised Manual*. Port Huron, MI: Sigma Assessment Systems.
- Judge, T., Bono, J.E., Ilies, R. & Gerhardt, M.W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87, 765-780.
- Judge, T. A., Heller, D. & Mount, M. K. (2002). Five-Factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, 87, 530-541.
- McCrae, R.R. & Costa, P.T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51, 882-888.
- McFarland, L.A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality tests scores. *International Journal of Selection and Assessment*, 11, 265-276.
- Meade, A.W. (2004) Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, 77, 531-552.
- Meehl, P.E. & Hathaway, S.R. (1946). The K factor as a suppressor variable in the MMPI. *Journal of Applied Psychology*, 30, 525-564.
- Mount, M. K. & Barrick, M. R. (1995). The Big Five personality dimensions: implications for research and practice in human resources management. In K.M. Rowland & G. Ferris (Eds.), *Research in personnel and human resources management*, vol. 13 (pp 153-200). Greenwich, CT: JAI Press.
- Mueller-Hanson, R., Heggstad, E.D. & Thornton, G.C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, 88, 348-355.
- Ones, D.S. & Viswesvaran, C. (1998a). Integrity testing in organizations. In R.W. Griffin, A. O'Leary-Kelly & J. Collins (eds). *Dysfunctional behaviour in organizations: Vol. 2. Nonviolent behaviours in organizations* (pp. 243-276). Greenwich, CT: JAI Press.
- Ones, D.S. & Viswesvaran, C. (1998b). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, 11, 245-269.
- Ones, D.S. & Viswesvaran, C. (2001a). Personality at Work: criterion-focused occupational personality scales (COPS) used in personnel selection. In B. Roberts & R.T. Hogan (Eds), *Applied personality psychology* (pp. 63-92). Washington, DC: American Psychological Association.
- Ones, D.S. & Viswesvaran, C. (2001b). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, 9, 31-39.
- Ones, D.S., Viswesvaran, C. & Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660-679.
- Ones, D.S., Viswesvaran, C. & Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validi-



- ties: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology (Monograph)*, 78, 679-703.
- Ones, D.S., Viswesvaran, C. & Schmidt, F.L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal of Personality*, 17, 19-38.
- Paulhus, D.L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.
- Paulhus, D.L. (2002). Socially desirable responding: The evolution of a construct. In H.I. Braun & D.N. Jackson (Eds.). *Role of constructs in psychological and educational measurement* (pp 49-69). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rolland, J.P. & Mogenet, J.L. (2001). *Système de description en cinq dimensions (D5D). Manuel réservé aux psychologues*. Paris: Les Editions du Centre de Psychologie Appliquée.
- Rosse, J.G., Stecher, M.D., Miller, J.L. & Levin, R.A. (1998). The impact response distortion on pre-employment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634-644.
- Rothman, S., Meining, D. & Barrick, M.R. (2002). *Predicting performance using personality measures in South Africa*. Symposium on Personality at work in a cross-cultural context (D.S. Ones, Chair), Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada, April.
- Salgado, J.F. (1996) Análisis exploratorio y confirmatorio del inventario de personalidad de cinco factores (IP/5F). *Psicológica*, 17, 353-366.
- Salgado, J.F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30-43.
- Salgado, J.F. (1998). The Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance*, 11, 271-288.
- Salgado, J.F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10, 117-125.
- Salgado, J.F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, 76, 323-346.
- Salgado, J.F. (2004). *Moderator Effects of Job Complexity on the Big Five Validity*. Poster presented at the Conference of the Society for Industrial and Organizational Psychology, Chicago, Illinois.
- Salgado, J.F. & De Fruyt, F. (2005). Personality in personnel selection. In A. Evers, O. Schmit-Voskuyl & N. Anderson (Eds). *Handbook of personnel selection*. Oxford, UK: Blackwell.
- Salgado, J.F., Remeseiro, C. & Iglesias, M. (1996). Personality and Test-Taking Motivation. *Psicothema*, 8, 553-562.
- Salgado, J.F. & Moscoso (2000). Autoeficacia y criterios organizacionales de desempeño. *Apuntes de Psicología*, 18, 179-191.
- Seisdedos, N. (1988). Selección, cuestionarios y distorsión motivacional (actitud inteligente de adaptación). *Revista de Psicología del Trabajo y las Organizaciones*, 4, 3-33.
- SHL (1999). *OPQ32 Manual and User's Guide*. Surrey: SHL.
- Van Rooy, D. L. & Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior*, 65, 71-95.
- Viswesvaran, C., Ones, D.S. & Hough, L.M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings? *International Journal of Selection and Assessment*, 9, 277-289.
- Yoo, T.Y. & Ming, B.M. (2002). *A meta-analysis of the Big Five and performance in Korea*. Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada, April.

