A Theoretical Model of Psychometric Effects of Faking on Assessment Procedures: Empirical findings and implications for personality at work

Jesús F. Salgado

Faculty of Labor Relations, University of Santiago de Compostela, Campus Vida, 15782, Santiago de Compostela, Spain. jesus.salgado@usc.es

This article proposes a theoretical model for explaining the psychometric effects of faking on assessment procedures (e.g., biodata, interviews, assessment center, personality inventories, and self-reported measures). The model hypothesizes that faking is a phenomenon of homogenization of scores, consisting of a double mechanism that increases the mean, on one hand, and decreases the standard deviation of distributions of scores, on the other. Subsequently, this affects the covariance, reliability, and validity of assessment procedures negatively. The model predicts that a mean ratio (faking ratio) greater than 1 and a coefficient of homogeneity *u* smaller than 1 characterizes faking. Meta-analysis was used to test several predictions of the theoretical model in the case of the personality measures. A database of 46 independent studies was created. All the studies used the NEO-PI-R for assessing the Big Five and their facets. The general pattern of data fully supported the model predictions. Implications for personnel assessment and, particularly, for personality assessment are discussed.

1. Introduction

Faking is a pervasive phenomenon that affects assessment procedures (e.g., personality test, biodata, assessment centers, and interviews) in applied contexts such as personnel selection and work promotion processes, academic decisions, and clinical and forensic settings. It is a bias consisting of a voluntary distortion of individual responses to assessment procedures (Furnham, 1986, 1990; Levashina & Campion, 2007). The objective of faking is to adapt the individual's image to the factors that the individual believes will produce advantages and benefits during the assessment process. Some researchers have stressed that this is an intelligent adaptation of individuals to the pressures and demands of the evaluative environment (Bangerter, Roulin, & Koning, 2012; Seisdedos, 1993).

The distortion of responses can be either positive or negative. The first is usually referred to as faking good and the second as faking bad. Faking good is typical in personnel selection and academic decisions where some individuals try to show higher scores in assessment procedures (e.g., personality inventories, interviews, biodata, assessment centers, interest forms, and other self-reported measures) they assume to be relevant for the job and the decision-making process. Faking bad is typical in clinical and forensic assessment. For example, some individuals may fake some symptoms and syndromes because they think that this simulation will support their case in court. From the point of view of the effects on the assessment process, both faking good and faking bad would produce similar effects on the psychometric properties (e.g., reliability and validity) of the assessment procedures (e.g., personality inventories, biodata, and interviews). In the case of positively scaled dimensions, the main difference between faking good and faking bad would be that faking good would produce an increment in the mean of some factors and faking bad would produce a decrease in the mean of the same factors. For instance, regarding emotional stability, individuals would increase positive responses in selection processes while negative

International Journal of Selection and Assessment

9600 Garsington Road, Oxford, OX4 2DQ, UK and 350 Main St., Malden, MA, 02148, USA

responses would be more frequent in a forensic assessment. In the case of negatively scaled dimensions, the main difference would be just the opposite.

This article is divided in two main sections. In the first section, after reviewing previous theoretical explanations of faking, I propose a new theoretical model of the effects of faking on assessment procedures. Specifically, the model explains how faking affects covariance, correlation, standard error, reliability, and validity of the assessment procedures that can be voluntarily distorted by individuals. The proposed model is general, in the sense that the same explanatory mechanisms are valid for all the assessment procedures that can be affected by faking such as, for instance, biodata, selection interviews (e.g., situational interviews), assessment centers, vocational interest questionnaires, and personality inventories. The first section ends with the revision of the empirical research on the faking effects in personality at work. Personality was chosen because of the special status of personality inventories with regard to faking research, First, personality inventories are much more used than biodata, assessment center, and situational interviews in personnel selection. Second, the effects of faking on personality inventories have been much more researched than on other selection procedures (e.g., interview, biodata, and assessment centers). Third, the theoretical foundations of the five-factor model (FFM) of personality are stronger than any other procedure affected by faking. For instance, the meaning of the personality dimensions and facets is much clearer than the meaning of what is measured by interviews, biodata, or assessment centers (Collins et al., 2003; Rothstein et al., 1990; Salgado & Moscoso, 2002). The second part of the article is devoted to the meta-analytic examination of some predictions of the model in the case of personality assessment in personnel selection.

2. Theoretical explanations of faking and its effects

In recent years, several models have been proposed for explaining both the antecedents (psychological process) and the consequences (psychometric effects) of faking. Snell, Sydell, and Lueke (1999) and McFarland and Ryan (2000, 2001) proposed the first two models of the psychological process of faking. The model proposed by Snell et al. (1999) includes two main elements: individual differences and situational characteristics. The variables of individual differences include ability to fake (e.g., general mental ability, experience) and motivation to fake (age, gender, personality). The situational variables include social desirability and importance of the outcome, and restraints. The model suggested by McFarland and Ryan (2000) is based on planned action theory (Ajzen, 1991), and it suggests that faking behavior is mainly dependent on the intention to fake, which, in turn, depends on the individual's attitudes toward faking, on subjective norms and on situational constraints. Both models have received some empirical support (e.g., Grieve & McSwiggan, 2014). Based on these two models, Mueller-Hanson, Heggestad, and Thornton (2006) developed a third model of faking antecedents, which includes dispositional and attitudinal antecedents. These variables are antecedents of intentions to fake, which in turn are the directly influence faking. Mueller-Hanson et al.'s model includes five antecedents: emotional stability, conscientiousness, ability to fake, willingness to fake, and perceptions of situation. More recently, Goffin and Boyd (2009), Griffith, Lee, Peterson, and Zickar (2011), and Marcus (2009) have proposed three additional psychosocial models of the antecedents of faking process.

With regard to the consequences of faking, two particularly relevant models were suggested by Heggestad, George, and Reeve (2006) and by Ziegler and Buehner (2009). Both models are based on the assumption that faking represents a source of measurement error in personality scores. Heggestad et al. (2006) suggested that faking introduces transient error to observed scores, which reflects the interaction between persons and situations. According to Heggestad et al. (2006), faking, as a transient error, would reflect variance in individual scores across situations due to sources unrelated to the construct intended to be measured. Using a two-month-apart repeated measure design, Heggestad et al. (2006) found larger transient error in faking response conditions than in honest response conditions. These researchers also found that the internal consistency estimates substantially overestimated the reliability of faked scores. This last finding suggests that coefficients of equivalence (CE), such as alpha, Spearman-Brown, and KR-20 coefficients might be less useful for estimating the reliability of assessment procedures responded to in faking-motivating situations. On average, CE overestimated the coefficient of equivalence and stability (CES) (Schmidt, Le, & Ilies, 2003) by 27.5% under faking conditions, while the overestimation was 9.7% under honest conditions. The transient error variance was .08 for honest conditions and 0.20 for faking.

Ziegler and Buehner (2009) proposed an approach based in the assumption that faking represents a source of spurious measurement error caused by the interaction of personality and situational variables. As a spurious error measurement, faking produces effects such as common method variance, increased correlations between variables, and lower criterion-related validity. The results of Ziegler and Buehner's (2009) study supported their hypotheses and showed that the effects of faking on means and covariance structure can be reversed. They also found that the variances of trait and faking might be separated.

In spite of the relevant contributions of these two models for the explanations of some effects of faking (e.g., on measurement error and criterion validity), they do not explain other important faking effects such as the variability in the number of personality factors after faking (an equal, greater, or smaller number of personality factors) or the possible change of the sign of the validity coefficients after faking (Tauriz, 2011). They do not include the potential effects on the *SD* of personality variables. Taking all of this into account, this article extends the previous models and presents a more comprehensive theoretical account of the effects of faking. It takes into account not only the faking effects on the mean but also on the standard deviation, covariance, reliability, and validity of assessment procedures, such as personality inventories. This new model is described in the next section.

3. A new theoretical model of faking effects

The present model is based on the idea that faking is a phenomenon of homogenization of samples (artificial self-selection) that, on average, causes individuals to appear more similar to each other than they really are. In other words, due to faking, smaller differences are observed in assessment procedures in situations which activate favorable self-presentations (e.g., personnel selection, forensic assessment, and academic decisions). According to the model, faking operates through a double mechanism. On the one hand, faking has an important influence on the increment or decrement of the mean of measured variables (e.g., personality dimensions, biodata, interviews), depending on whether it is faking good or faking bad. This influence can be calculated as a faking ratio (FR) by dividing the mean under faking conditions by the mean in honest conditions. In the case of positively scaled variables, the FR value should be greater than 1 if faking good exists and the FR would be smaller than 1 if faking bad exists. The FR will be 1 if faking does not exist at all. This effect of increment or decrement of the mean has been well established, although not as a FR value but as Cohen's *d*, as was shown in the empirical research review section (e.g., Birkeland, Manson, Kissamore, Brannick, & Smith, 2006; Ones & Viswesvaran, 1999).

On the other hand, faking has a remarkable influence on the variance (and *SD*) of the measures. This second effect of faking reduces the standard deviation of measures (e.g., personality dimensions). Therefore, the comparison of the standard deviations of a faking group and of an honest group (SD_f/SD_h) should produce a value smaller than 1. To the best of my knowledge, this second effect of faking has not been systematically examined until now. However, the examination of some reported findings support the possibility that this may be true. For example, an examination of Tables 1–4 of Ellignson, Smith, and Sackett (2001) article reveals that out of 62 SD_f/SD_h ratios, in 56 (90.2%) the value is smaller than 1, in 3 (4.8%) the value is 1, and in 3 (4.8%) the value is slightly larger than 1. In other words, under faking conditions, the increment (or decrement) of the mean scores is simultaneously accompanied by a reduction in the size of the standard deviation, which revels a restriction of the range scores of the distribution. Such a restriction could be produced in both extremes of the distribution of personality scores, depending on the type of faking (good or bad). The restricted distribution indicates that faking could be an artificial self-selection process.

An important consequence of this line of reasoning is that faking, through the artificial alteration of the mean and standard deviation of the original distribution, affects all coefficients based on these two statistics, mainly the following four: (a) covariance, (b) reliability, (c) standard error, and (d) correlation (e.g., validity). The influence of both natural selection (i.e., due to changes in the environment) and artificial selection (i.e., due to the activities of humans) on mean, variance, and correlation was first considered by Karl Pearson (1903), and this faking model is inspired by some of his theoretical and mathematical contributions.

The effect of faking on the mean occurs by increasing (or decreasing) its value, depending on the self-presentation and impression management objectives of the individuals (Levashina & Campion, 2007). However, in the case of the standard deviation, the effect of faking will always result in a reduction of the differences between individuals, producing homogeneity through the restriction of the range of the scores. Typically, the degree of range restriction is estimated with the coefficient of heterogeneity *U* (Burt, 1943; Pearson, 1903), which is obtained as the ratio between the *SD* of the nonrestricted group and the *sd* of the restricted group. Therefore

$$U=\frac{SD}{sd} \tag{1}$$

The inverse of coefficient *U* can be called coefficient of homogeneity and it is typically represented as *u*. There are three additional ways to obtain the coefficient of heterogeneity (and homogeneity). The first is based on the selection ratio. Kelley (1947) and Schmidt, Hunter, and Urry (1976) derived the same formula, independently each other, for obtaining *u* from the selection ratio. The *U* coefficient (and the *u* coefficient) can also be obtained from the reliability coefficient of the unrestricted group and the reliability coefficient of the restricted group. Based on Kelley's (1921) formula, $SD\sqrt{1-R_{xx}}=sd\sqrt{1-r_{xx}}$ and, by carrying out some algebraic transformations, it is possible to obtain the following two equations:

$$\frac{SD}{sd}=\frac{\sqrt{1-r_{xx}}}{\sqrt{1-R_{xx}}}=U \tag{2}$$

and

$$\frac{sd}{SD}=\frac{\sqrt{1-R_{xx}}}{\sqrt{1-r_{xx}}}=u \tag{3}$$

where
R_{xx} = reliability in the nonrestricted group
r_{xx} = reliability in the restricted group.

Recently, Hunter, Schmidt, and Le (2006, p. 601) developed a new formula for estimating the coefficient of homogeneity u under indirect range restriction in the predictor, which also connects homogeneity and reliability. The formula is based on the observed u value. The formula is as follows:

$$u_T = \sqrt{\frac{u_x^2 - (1 - r_{xxa})}{r_{xxa}}} \tag{4}$$

where
u_T= the true value of u under indirect range restriction
u_x= the observed
ur_{xxa}= the reliability of predictor in the applicant population (unrestricted group)

Both $1\text{-}R_{xx}$ and $1\text{-}r_{xx}$ terms are error variance. Kelley (1921), Otis (1922), Brogden (1944), Hunter et al. (2006), and many other researchers, have suggested that it is reasonable to assume that the error variance is equal for the true scores in the unrestricted and restricted groups. They also suggested that it can also be assumed that it is uncorrelated with all selections which can produce changes in the variance of Y. It is relevant to note that these formulas show that the homogeneity (range restriction) of the sample scores necessarily results in lower reliability and that the heterogeneity (range enhance) results in a larger reliability. Consequently, the effects of faking on the reliability coefficients found in empirical research (e.g., Heggestad et al., 2006) could be explained through the homogeneity mechanism of faking.

With regard to covariance, faking has two effects through its influence on the mean and SD of the faked variable. Firstly, it produces a smaller value of covariance, as consequence of the increment of the predictor mean. The second effect on covariance is that the sign may be changed from positive to negative (and vice versa), depending on the indirect effects of X on the criterion (covariate) mean. The covariance formula is:

$$COVx = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{N - 1} \tag{5}$$

Faking affects the first term of the numerator $(X - \bar{X})$ directly and the second term of the numerator $(Y - \bar{Y})$ indirectly. It is easy to see that as N is constant, any change in the mean of X will change the covariance value. The change of the covariance sign depends on whether the indirect effects on Y are univariate or multivariate (Thomson, 1938, 1939, 1944; Thurstone, 1945, 1965). If the indirect effect is univariate, no change will be produced in the sign. If the indirect effects are multivariate, then it is possible that the covariance will change the sign. For example, Thorndike (1947; see also Ree, Carreta, Earles, & Albert, 1994) reported a change of the sign of the validity coefficients due to indirect range restriction. These effects on the covariance have dramatic consequences for estimating correlation coefficients, which are the basis on which to calculate other coefficients such as reliability and validity, and to carry out factor analyses. Furthermore, the covariance sign is what determines the sign of the correlation coefficient.

The effect of faking on the correlation coefficient is also determined by the effects of faking on the covariance of X and Y, and the effects of faking on the standard deviation of X and Y. Pearson (1903; see also Hunter et al., 2006) demonstrated that a direct restriction on X is always accompanied by an indirect restriction on Y, although to a smaller degree. Consequently, the restriction of SD_x is accompanied by a restriction in SD_y. The formula of the correlation coefficient is:

$$r_{xy} = \frac{COVxy}{\sigma_x \sigma_y} \tag{6}$$

Following this rationale, it can be seen that the reduction in covariance size and the potential sign change determines the size and the sign of the correlation coefficient. Additionally, the correlation coefficient is affected by the size of the standard deviation of X and Y. As the model assumes that faking is a process of homogeneity, the greater the intentional distortion, the smaller SD_x, but the reduction of changes in SD_y would be comparatively less important than the changes in SD_x. This combination of reduced SD_x and SD_y, together with a smaller covariance, would produce a smaller validity coefficient. The modification of the correlation coefficient, by reducing it and changing its sign, has dramatic consequences on the criterion-related validity but also on factor analysis and on covariance structure analysis, for instance.

With regard to construct validity, it is important to take into account, as was aforementioned, that faking can produce univariate or sequential range restriction (selection) depending on the number of variables (e.g., personality dimensions) that are affected (Frank L. Schmidt, personal communication, 3/4/2014).[1] The first case is when range restriction occurs on one variable. When several variables are affected by faking, what is produced in this situation is a sequential univariate indirect range restriction, on one measure after another, which will produce the same effects as multivariate range restriction. Consequently, the effects of faking on the construct validity vary depending on the type of range restriction (univariate or sequential). Typically, the construct validity of assessment procedures (e.g., personality inventories) is examined using factor analysis. In a series of papers and two books, Godfrey Thomson and his students (Lawley, 1943; Ledermann, 1938; Thomson, 1938, 1939, 1944; Thomson & Ledermann, 1939) and Thurstone (1945, 1965) demonstrated mathematically that both the univariate and multivariate selection determines the results of a factor analysis and that the effects were different for the two types of selection. Univariate selection reduces common variance and increases specific variance of all variables correlated with the restricted variable. In fact, the increment of the specific variance is equal to the ratio of q_1^2/u^2, where q_1^2 is $1-u^2$. Hence, in the case of the univariate selection, the number of common factors in the tests would remain the same, but their loadings or saturations would decrease.

The effects of sequential (multivariate) selection are different. Sequential selection can produce additional factors, depending on the number of variables in which the selection occurs. Furthermore, the new factors would be uncorrelated with the older factors. Thomson and Lederman (1939) showed that the number of induced factors would be small or equal to

$$\frac{2p+1-\sqrt{8p+1}}{2} \quad (7)$$

where p is the number of original factors. Conversely, Thurstone (1935, p. 76) suggested that one of the most important sources of induced factors was what he called experimental dependence. Faking is a specific source of experimental dependence. Once again, considering the formula of covariance and the formula of the correlation coefficient, it is possible to figure out the effect of the sequential selection. If faking affects two or more variables (e.g., X_1 and X_2), then the covariance between X_1 and X_2 will be greater, but their standard deviations will be slightly smaller. Consequently, the correlation between X_1 and X_2 will be larger. If this phenomenon occurs for multiple variables, a new (spurious) factor arises.

The demonstrations by Thomson and his students are especially relevant with regard to some of the contradictory findings on the effects of faking on the structure of personality inventories. For example, they permit a mathematical explanation of Schmit and Ryan's (1993) results, concerning the factor of 'ideal employee'. Due to the fact that Schmit and Ryan (1993) used the NEO-PI with the five original factors, and that the range restriction occurred in more than one variable, the sequential selection created by faking would produce an extra factor, in line with what is suggested by formula 7. Thomson's demonstrations also explain the results provided by Ellingson et al. (2001), Michaelis and Eysenck (1971), and Van Iddekinge et al. (2005) equally well, under the assumption that faking represents a mechanism of range restriction. Thomson's demonstrations are also in accordance with the finding of Heggestad et al. (2006) that the spurious variance caused by faking produces larger intercorrelations between the scales, which might produce a common method factor.

In summary, the model of faking effects proposed here is based on and extends the previous psychometric models by Heggestad et al. (2006) and Ziegler et al. (2010). It relies on the assumption that faking influences the scores of distorted variables through two important and simultaneous artifactual effects. The first artifactual effect is on the mean of the distribution of scores, causing an increment (faking good) or a decrement (faking bad) of this statistic. The second artifactual effect is on the variance (and *SD*) of the measures. In other words, the increment (or decrement) in the scores of voluntarily faked assessment procedures, which has repeatedly been found in empirical research, is simultaneously accompanied by a reduction in the size of the standard deviation, resulting in a restriction of the distribution of scores. The main consequence of the artifactual increment (or decrement) of the mean is a degree of common method variance, which becomes larger when faking is more extreme. Thus, the ratio between the mean in honest (control) conditions and the mean in faking conditions would be greater than 1 (in faking good conditions). Conversely, the restriction would take place at both extremes of the distribution, but it would be larger at the higher extreme (in faking good conditions). Therefore, the comparison of the standard deviations between an honest (nonrestricted) group and a faking-good group (restricted) would produce a ratio greater than 1. As previously demonstrated, this artifactual alteration of the mean and *SD* of the original distribution produced by faking will have consequences over all statistics and coefficients based on them, such as the covariance, reliability, standard error of measurement, and construct and criterion-oriented validity.

4. Empirical research on faking effects in personality at work: A summary

Personality inventories began to be used as assessment procedures almost a 100 years ago. Since then, faking has recurrently been considered a problem that reduces their utility (Ellis, 1946; Guilford, 1954; Meehl & Hathaway, 1946; Morgeson et al., 2007a,b; Murphy & Dzjeweczysnki, 2005; Zickar & Gibby, 2006) and a recent study on faking prevalence showed that, depending on the confidence interval used, between 30% to 50% of applicants fake on personality inventories (Griffin, Chmielowski, & Yoshita, 2007).

The vast majority of research on the effects of faking on personality measures has focused on four major issues: (a) the increment of the scores, (b) the change of the rank-order of the individuals, (c) the modification of the inventory factorial structure (construct validity), (d) the lowering of criterion-related validity, and (e) the attenuation of reliability. A summary of this research follows.

Many individual studies, both in experimental settings and in occupational, clinical, and educational settings, have found that when the assessees voluntarily fake or distort their responses to the items, this distortion results in a score higher or lower in some personality factors (e.g., Anderson & Sleap, 2004; Rosse, Stecher, Miller, & Levi, 1998; Scarpello, Ledvinka, & Bergman, 1995; Schmit & Ryan, 1993; Salgado, 2005; Van Hooft & Born, 2012). Viswesvaran and Ones (1999; see also Ones & Viswesvaran, 1998) conducted a meta-analysis in which they examined the differences between the effects of faking and honest-response conditions on the Big Five personality dimensions. They found that the effect size of faking varies depending on study design. More specifically, the effect size of faking was $d = 0.60$ on average for between-subject designs and $d = 0.71$ when the design is within-subject. Consequently, Viswesvaran and Ones (1999) concluded that personality measures can be faked when individuals are motivated to do so. This meta-analysis was recently supplemented by another meta-analysis carried out Birkeland et al. (2006), in which they examined the effects of faking on the Big Five scores in applicant and nonapplicant samples. Birkeland et al. (2006) found that applicants and nonapplicants mainly differ in the factors of emotional stability and conscientiousness, with effect sizes (Cohen's *d*) of 0.44 and 0.46, respectively.

A second effect of faking is that it can potentially change the rank-order of the individuals when a top-down ranking is used. As Morgeson et al. (2007a,b, p. 685) suggested, '*faking might do more than simply elevate test scores; it might change the rank-order of examinees, and it might lead to different rank orders in different situations even if the pool of examinees is kept constant*'. Ellingson, Sackett, and Hough (1999) examined this point and found that the honest individuals holding the top ranks were displaced when other individuals fake. They also found that the application of faking corrections did not reestablish the rakings in all cases. Similar results were found by Christiansen, Goffin, Johnson, and Rothstein (1994), Douglas, McDaniel and Snell (1996), Rosse et al. (1998), and Zickar (2000; Zickar, Rosse, & Levin, 1996). More recently, Griffin et al. (2007) found dramatic changes in the rank order of individuals in a sample of applicants that were asked to respond to measures of personality in two additional conditions, honest and faking.

With regard to the effects of faking on the construct validity of personality inventories, the results are mixed as a number of studies using factor analysis methodology have reached contradictory conclusions. It was found that under faking instructions, personality inventories may produce an equal, larger, or smaller number of factors than under honest instructions. For example, Michaelis and Eysenck (1971) found that the results are exactly the same under faking conditions and with honest answers in a study using the Eysenck Personality Questionnaire. Similar findings were reported more recently by Smith, Hanges, and Dickson (2001). These researchers used the Hogan Personality Inventory (HPI, Hogan & Hogan, 1995) in three large samples of applicants, incumbents, and students consisting of 2,500 individuals each. Smith et al. (2001) found exactly the Big Five in the three cases. Similarly, a study carried out by Ellingson et al. (2001), consisting of four very large samples of applicants and incumbents, who responded to the ABLE, CPI, 16PF, and HPI inventories respectively, found that confirmatory factor analyses demonstrated that the factor structures were equivalent, although not parallel, for the four inventories. This means that the scales loaded on the same factor but that some loadings were different across the scales and factor analyses. However, contrary to the findings mentioned above, Schmit and Ryan (1993) and Cellar, Miller, Doverspike, and Klawsky (1996) found that faking produced a distorted structure in the NEO PI under faking instructions, showing six factors when only five were expected. Finally, Ellingson et al. (1999), Frei et al. (1997), Van Iddekinge, Raymark, and Roth (2005) found that the number of factors was reduced under instructions to fake.

In connection with criterion validity, the meta-analysis by Ones, Viswesvaran, and Reiss (1996) suggested that evidence of validity is not affected by faking, when it is assessed with social desirability responding scales (SDR). Globally, the criterion validity remained essentially the same when the effects of SDR were partialled out. A contrary view about the effects of faking on the criterion-related validity of personality inventories was offered by Hough (1998). Based on a review of the empirical evidence available, Hough concluded that the majority of the studies supported the hypothesis that faking seriously lowered criterion validity (one study was the exception) and, more importantly, that validity decreases as the amount of intentional distortion increases.

With respect to the effect of faking on the reliability of personality inventories, few studies have been conducted to examine this point. However, some articles include information about the reliability in honest and faking conditions, which allow us to examine the potential effects of faking on reliability coefficients. For example, Heggestad et al. (2006) found a larger transient error in faked inventories and, consequently, smaller alpha coefficients. Similarly, Van Iddekinge et al. (2005) found smaller reliability coefficients under faking conditions than in honest conditions. More recently, Van Hooft and Born (2012) found the same effect on reliability using a different personality inventory than the one used in the previous studies. In this laboratory study, participants were instructed to respond twice to a personality inventory based on the FFM, the first time honestly and the second faking good in an attempt to appear the most suitable job applicant. The results showed that internal consistency (Cronbach's alpha) was smaller for extroversion (0.93 vs. 0.84), conscientiousness (0.90 vs. 0.85), emotional stability, (0.91 vs.

0.84), and autonomy (0.91 vs. 0.83) and was slightly larger for agreeableness (0.81 vs. 0.83) and integrity (0.71 vs. 0.72) under faking conditions. Consequently, the studies mentioned above suggest that faking also affects the estimates of reliability, reducing their size in comparison with the coefficients obtained under honest conditions.

In summary, research has showed that when individuals are instructed or are motivated to distort a personality inventory they can do so in accordance with their instructions (faking good or faking bad). The majority of studies on faking have been centered (a) on the effects on the mean of personality measures, (b) the change in rank-order, and (c) the potential effects on the criterion-oriented validity. In general, research has shown that faking good resulted in higher scores, and faking bad in lower scores in the selected personality measures positively related to the criterion of interest (e.g., job performance). In the case of faking good, research on personnel selection also showed that there are differences among samples of nonapplicants, applicants, and incumbents.

However, the potential effects of faking go beyond this, as faking can negatively affect other statistics such as the standard deviation of personality measures, their reliability, and their construct validity. The theoretical explanations that have been proposed on the antecedents and effects of faking are reviewed in the next section.

5. Empirical examination of the model

The credit of any theoretical model relies on the degree in which their main premises were empirically tested and supported. In the present case, it is necessary to demonstrate that: (1) faking produces a mean FR greater than 1 for the groups that respond to personality inventories under conditions that motivate them to fake (e.g., applicants, motivated-to-fake incumbents, and experimental groups); (2) the coefficient of homogeneity u is smaller than 1 for those groups in comparison with nonmotivated-to-fake groups (e.g., normative groups, honest-conditions groups), and (3) faking produces smaller reliability coefficients, which, together with range restriction, constitute the two main artifactual errors affecting validity.

For the reasons mentioned at the beginning of the article, the Big Five personality dimensions have been chosen to empirically test the model in a personnel selection context. Due to the fact that many single studies already exist which report data concerning the effects of faking, meta-analysis seems to be the most appropriate method to demonstrate the new explanatory model. Consequently, a meta-analysis of the effects of faking on the mean, standard deviation, and reliability of the Big Five and their facets was conducted. As faking good is typical in personnel selection, only studies concerning the first type of faking were selected for the meta-analysis. To avoid potential contaminating effects due to variability in construct measurement (Schmidt & Hunter, 2014, p. 94) which could arise from the use of different personality inventories, only studies using the NEO-PI and NEO-PI-R (Costa & McCrae, 1992) were meta-analyzed. This inventory was selected because it is probably the most popular of the personality inventories based on the FFM. Furthermore, it contains a number of scales of facets, which also allows the examination of the faking effects both at the level of dimensions and at that of facets.

In the present case, three types of studies were included in the meta-analysis. The first two types included real-world applicants and real-world incumbents. A third type of studies was also meta-analyzed: experimental studies using within-subjects designs. These studies included students and, for this reason, they were not collapsed with the other samples; that is, applicants, incumbents, and students were separately meta-analyzed. An anonymous reviewer suggested that the faking effects apply to the between-subjects designs, too. Of course, the faking effects predicted by the model apply to both between-subjects designs and within-subjects designs. However, in this point, the results of the meta-analysis by Viswesvaran and Ones (1999) are relevant. Viswesvaran and Ones (1999) compared the fakability estimates of within-subjects designs and between-subjects designs. They found that the fakability estimates were larger for the within-subjects designs than for between-subjects designs). Moreover, Viswesvaran and Ones (1999) suggested that within-subjects designs produce more accurate estimates of the effects of faking, among other things, because the between-subjects design assumes that the effect size is the same for all individuals, and that the interaction Instruction Set × Individual Propensity to Fake is nonexistent. Finally, Visweswaran and Ones (1999) suggested that the sample equivalence of groups assumed in the between-subjects designs seems questionable in most studies. Consequently, the statistical power of the within-subjects designs is larger than that of the between-subjects designs.

Based on the model of the faking effects proposed here, the following five hypotheses are advanced.

Hypothesis 1: Individuals motivated to fake good show personality distributions characterized by higher means in comparison with individuals less motivated to fake. Therefore,

Hypothesis 1a: Applicants and participants in within-subject designs will show an average FR greater than 1. This average FR will be larger than the FR of incumbents groups and honest groups. Additionally, the estimates of standardized mean differences (Cohen's d) will show that applicants and individuals induced to fake in within-subject designs score higher than incumbents groups and honest groups.

Hypothesis 1b: The motivated-to-fake incumbents will show an average FR larger than the average FR of the nonmotivated incumbents. The standardized mean differences will show that the motivated-to-fake incumbents score higher than the nonmotivated incumbents.

Hypothesis 2: Individuals motivated-to-fake good show personality distributions characterized by smaller *SDs* in comparison with samples of individuals less motivated to fake. Therefore,

Hypothesis 2a: Applicants and individuals induced to fake good in within-subject designs will show coefficients of homogeneity *u* smaller than 1. Additionally, their *u* values will be smaller than the *u* of the incumbents groups and noninduced-to-fake groups.

Hypothesis 2b: The motivated-to-fake incumbents will show smaller coefficients of homogeneity than the nonmotivated incumbents.

Hypothesis 3: The reliability of the scales for measuring the Big Five and their facets is negatively affected by faking. Therefore, the reliability estimates calculated with applicants and individuals induced-to-fake good will be smaller than the reliability estimates found in samples of nonmotivated-to-fake incumbents, volunteers, and normative groups.

6. Method

6.1. Literature search

An extensive literature search was conducted using several approaches and databases to identify both published and unpublished studies that have examined faking using the NEO-PI-R. First, the databases PsycInfo, EBSCO, ScienceDirect, Wiley Online Library, Sage, and Google Scholar were searched to identify studies into the NEO PI-R and faking. Several keywords were used for the computer-based literature search, for example '*NEO PI-R*', '*faking*', '*personality validity*', '*applicants*', '*incumbents*', '*norms*', '*NEO PI-R reliability*', '*motivational distortion*'. Second, a manual search was conducted article-by-article in the following journals since 1992: *Journal of Applied Psychology, International Journal of Selection and Assessment, Personnel Psychology, Journal of Work and Organizational Psychology, Journal of Occupational and Organizational Psychology, Applied Psychology-An International Review, and Journal of Personnel Psychology.* Third, the reference sections of several past meta-analyses (e.g., Barrick & Mount, 1991; Birkeland et al., 2006; Hurtz & Donovan, 2000; Salgado, 1997, 1998, 2003; Tett, Rothstein, & Jackson, 1991; Viswesvaran & Ones, 1999) were examined to identify articles not located in the two previous approaches. Finally, a number of researchers were contacted to obtain additional papers. The final database comprises 37 papers, including 46 independent samples.

6.2. Procedure

As with many other meta-analyses, three types of variables were used: dependent, grouping, and moderators. The dependent variables of interest for this meta-analysis were mean, standard deviation, and the reliability of the NEO PI-R scales (both for the dimensions and facets). The grouping variables were the Big Five personality dimensions and their respective facets. The sample type (applicants, nonmotivated incumbents, motivated incumbents, experimental faking groups, and experimental honest groups) was the variable used as a moderator of the size of the dependent variables.

For each study, the following information was recorded, if available: (a) sample demographics, such as sex, occupation, and education, (b) type of NEO PI-R version used (e.g., 1987, 1992), (c) sample type: applicant, incumbents (motivated and nonmotivated to fake), and experimental (honest-condition group, faking-condition group); (d) sample size, (e) mean of the dimensions and facets; (f), standard deviations of the scales and facets or, alternatively, the coefficient of homogeneity *u* when reported; (g) reliability coefficients of the NEO PI-R dimensions and facets obtained under faking and nonfaking conditions. A PhD. student and I recorded the information for each study independently. The degree of agreement between the two researchers was practically complete. The very minor disagreements were resolved by referring to the original study.

The next step was to calculate the FR, the standardized mean difference (Cohen's *d*), and the *u* value for each of the personality dimensions and facets reported in the studies. To calculate the FR, I proceed as follows: (1) In the case of the applicant samples, the mean was divided by the mean of the normative group reported in the study or, if this value was not available, was divided by the mean of the norms published in the NEO PI-R manuals (e.g., US manual, French manual, Spanish manual); (2) for the incumbent samples, the mean ratio was calculated in a similar way as for the applicant samples, but using the incumbent mean in the numerator; (3) for the experimental studies of faking, we used the data of the within-subject designs only, and the ratio was computed as the mean of the faking-condition group divided by the mean of the honest-condition group. A number of studies reported *T* values. In this case, we divide the reported mean by 50 (*T*-score mean).

As was done by Viswesvaran and Ones (1999) and Birkeland et al. (2006), in addition to FR, the standardized mean difference (Cohen's *d*) was also calculated as an estimate of effect size of faking. To calculate Cohen's *d*, the meta-analytic program software of Schmidt and Le (2014) was used. This program allows to estimate the sample size–weighted mean *d*, the sample size–weighted *SD* of the observed mean *d*, the *SD* predicted due to sampling error, and the residual *SD*. The residual *SD* is the square root of the difference between the observed and sampling error variance and is used to establish the confidence intervals around the observed mean. Finally, the 80% confidence interval was computed. This interval indicates the range within which 80% of the observed ds were found.

The use of two estimates of effect size in this research serves for different purposes. FR is a useful way for detecting the presence/absence of faking because values larger than 1 indicate faking and values smaller than 1 indicate absence of faking. However, a problem with the FR arises when the scores of personality scales are reported in z scores. In this case, the FR would give infinite, independently of the degree of faking. This is not a problem in the present study because the FR was always computed using raw scores or *T* scores. Standardized mean differences were calculated to compare the results of this research with of those of previous meta-analyses (e.g., Birkeland et al, 2006; Viswesvaran & Ones, 1999).

In the case of the coefficient of homogeneity *u*, the following procedure was carried out: (1) if the study reported the *u* values, these were used in the database; (2) for the within-subject studies, the standard deviation of the faking condition was divided by the standard deviation of the honest condition; (3) for the applicant samples, their standard deviations were divided by the standard deviation of the norm group when reported in the study or was divided by the standard deviation reported in the NEO PI-R manuals (e.g., Spanish manual for the Spanish samples, French manual for the French samples, etc); (4) for the incumbent samples, an identical procedure was followed as for the applicants samples; (5) for the studies that reported *T* values, the *SD* was divided by 10 (standard deviation of *T*-scores).

With regard to the reliability coefficients, the vast majority of the reported coefficients were Cronbach's alpha. Therefore, this coefficient was recorded for each sample when reported. The alpha values were recorded independently for the studies with samples of motivated-to-fake individuals (e.g., applicants) and individual nonmotivated-to fake (e.g., volunteers, nonmotivated incumbents) for both the Big Five dimensions and their facets.

To have a better knowledge of the effects of faking, in the case of the samples of incumbents, these were divided into two categories: (a) nonmotivated-to fake incumbents, and (b) motivated-to-fake incumbents. The first consisted of the samples in which (1) the participants were explicitly informed that the personality scores would not be used for making a decision, (2) they were asked to be as honest as possible, or (3) they were informed that their scores would serve only for job profiling purposes. The second category consisted of the samples in which (1) the participants were not explicitly informed about the organizational purposes of the personality scores, (2) the participants were informed that the scores would condition the final decision, or (3) the incumbents participate in formal training courses or organizational processes which required them to respond to a personality inventory.

6.3. Meta-analytic method

A 'bare-bones' meta-analysis was carried out in this investigation (Hunter & Schmidt, 2004) to calculate the means of the three variables of interest (FR, coefficient of homogeneity *u*, and reliability), observed variances, sampling error variances, and variance corrected for sampling error. As the objective of this meta-analysis was to establish estimates of FR, *d*, *u*, and, r_{xx} for the Big Five personality factors and their facets as assessed by the NEO PI, under faking and nonfaking conditions, the sample-size weighted averages were computed on these three variables for the Big Five and also for their facets. This involved the computation of 90 individual meta-analyses.

One difference between the current meta-analysis and the usual meta-analysis of correlations and *d* is that ratios were also used as effect sizes. This type of effect size is less common in psychology than in other sciences (e.g., epidemiology, ecology, pharmacology) but they have been used in some psychological studies. For example, Ferguson and Bibby (2012) carried out a meta-analysis on openness to experience as a protective factor with respect to all-cause mortality, in which ratios were used as data. Hedges, Gurevitch, and Curtis, (1999) and Lajeneusse (2011), among others, examined the use of ratios as an effect index and illustrated how to conduct a meta-analysis of ratios. The formula of the variance of a ratio was given by Geary (1930), Fieller (1932), Dunlap and Silver (1986), Hedges et al. (1999), among others. It is:

$$V_R = R^2\left[\frac{CV_c^2}{n_c} + \frac{CV_e^2}{n_e}\right] \tag{8}$$

Where $V_R =$ variance of the ratio; $R^2 =$ squared ratio, $CV_c =$ coefficient of variation of the control group; $CV_e =$ coefficient of variation of the experimental group; and n_c and n_e are the respective sample size. This formula was used in the meta-analysis of FRs.

Results

7.1. Meta-analysis of the FR and Cohen's d of the Big Five and their facets across samples

The results of the meta-analysis of the FR and *d* for the Big Five personality dimensions appear in Table 1. From left to right, the first six columns represent the total

Table 1. Effect size (Cohen's *d*) and faking ratio (FR) of the three types of samples for the Big Five personality dimensions

Variable	*N*	*K*	*d*	SD_d	%VE	$80\%CI_d$	FR	S^2_{fr}	SD_{fr}	$80\%CI_{fr}$
Incumbents – all										
Emotional stability	5,467	18	0.67	.337	11	0.24–1.10	1.19	.0130	.115	1.04–1.34
Extraversion	5,798	20	0.84	.488	6	0.22–1.47	1.14	.0062	.079	1.04–1.24
Openness	5,645	20	0.38	.378	9	−0.10–0.87	1.06	.0040	.063	0.98–1.14
Agreeableness	5,467	18	−0.19	.302	13	−0.58–0.20	0.98	.0021	.047	0.92–1.04
Conscientiousness	5,896	21	0.45	.269	17	0.11–0.80	1.07	.0020	.044	1.01–1.13
Average			*0.43*	*.355*			*1.09*		*.070*	
Applicants										
Emotional stability	32,599	11	1.11	.232	3	0.81–1.41	1.28	.0051	.072	1.19–1.37
Extraversion	32,599	11	0.77	.173	5	0.55–1.00	1.13	.0025	.050	1.07–1.19
Openness	32,917	12	−0.14	.490	1	−0.77–0.48	0.99	.0057	.076	0.89–1.09
Agreeableness	31,203	10	0.56	.172	4	0.34–0.78	1.07	.0038	.019	1.05–1.09
Conscientiousness	33,002	13	1.18	.285	2	0.82–1.55	1.17	.0011	.033	1.13–1.21
Average			*0.70*	*.270*			*1.13*		*.050*	
Within-subject design										
Emotional stability	427	4	1.98	.243	49	1.66–2.29	1.69	.0015	.039	1.64–1.74
Extraversion	607	5	0.50	.214	43	0.23–0.79	1.07	.0009	.031	1.03–1.11
Openness	607	5	0.09	.016	99	0.07–0.11	1.01	.0008	.028	0.97–1.05
Agreeableness	607	5	0.94	.550	11	0.24–1.64	1.13	.0053	.073	1.04–1.22
Conscientiousness	607	5	1.52	.126	73	1.36–1.68	1.24	.0011	.033	1.20–1.28
Average			*1.01*	*.230*			*1.23*		*.041*	

Note: *N* = total sample; *K* = number of studies; *d* =sample size weighted mean effect size; SD_d = standard deviation in effect size remaining after sampling error variance is removed; %VE = % of observed variance accounted for sampling error; $80\%CI_d$ = 80% credibility interval of effect sizes; FR = weighted-sample mean; SD_{fr} = standard deviation of FR; $80\%CI_{fr}$ = 80% confidence interval for FR.

sample size, the number of independent studies, sample size weighted mean *d*, standard deviation in effect size (SD_d) remaining after sampling error variance is removed, percentage of observed variance accounted for sampling error, and 80% Credibility interval of effect sizes. The next four columns represent the sample size weighted mean FR, the variance of FR, the standard deviation of FR, and the 80% confidence interval around the mean of FR.

As can be seen, in the case of incumbents, the standardized mean differences were positive for emotional stability, extraversion, openness and conscientiousness, with *d* values ranging from 0.38 to 0.84. They were also positive for emotional stability, extraversion, agreeableness, and conscientiousness for applicants, with values values ranging from 0.56 to 1.18. In the case of the within-subject designs, all the Big Five showed positive standardized mean differences, with values ranging from .09 to 1.98. The magnitude of the *d* estimates was particularly remarkable for emotional stability, extraversion, and conscientiousness across the three types of samples. Globally, these findings agree with Viswesvaran and Ones' (1999) findings, with the exception of the results for openness to experience. The findings concur also with Birkeland et al.'s (2006) findings, although the magnitude of the effects sizes is very much larger in the present study.

With regard to FR estimates, in the case of the incumbent samples, FR varied from 0.98 to 1.19, with an average of 1.09. The largest FR was for emotional stability, followed by extroversion. In the case of applicants, FR varied from 0.99 to 1.28, with an average of 1.13. The largest FR was also for emotional stability, followed by conscientiousness. With regard to the within-subjects designs, FR varied from 1.01 to 1.69, with an average of 1.23. The largest FR was also for emotional stability, followed by conscientiousness. Therefore, as a whole, these results support *hypothesis 1a* as the three groups showed larger *d*-values and FRs than the normative or control groups. Furthermore, as expected applicants showed larger FR than incumbents.

The incumbent group was subdivided into two groups, one motivated to fake and the other nonmotivated to fake. The meta-analytic results for these two subgroups appear in Table 2. As can be seen, there are remarkable differences in the incumbents samples. The standardized mean differences were remarkably larger for motivated-to-fake incumbents than for nonmotivated incumbents. In fact, the *d*-values were very small for the five personality factors in the nonmotivated incumbents, with values ranging from −.05 to 0.18. FR ranged from 0.97 to 1.24 for the motivated incumbents and FR ranged from 1.00 to 1.04 for the nonmotivated group. The average FR is 1.1 and 1.03 for the motivated and nonmotivated group, respectively. The difference between these two ratios is statistically significant ($p < .001$, one-tail). This finding supports *hypothesis 1b*. Furthermore, the FR of nonmotivated incumbents is practically identical to that of the normative groups, and the FR of the motivated incumbents approached that of applicants. Therefore, this suggests that it is crucial to distinguish between incumbents who are motivated and nonmotivated to fake in future studies.

The meta-analysis of the faking effects for the facets of the Big Five is reported in Table 3. Not enough data were

Table 2. Effect size (Cohen's *d*) and faking ratio (FR) of motivated and non-motivated incumbents for the Big Five personality dimensions

Variable	N	K	d	SD_d	%VE	80%CI_d	FR	S^2_{fr}	SD_{fr}	80%CI_{fr}
Motivated incumbents										
Emotional stability	4,288	10	0.79	.162	28	0.58–0.99	1.24	.0057	.076	1.14–1.34
Extraversion	4,467	11	1.04	.320	10	0.63–1.49	1.17	.0026	.051	1.10–1.24
Openness	4,381	11	0.44	.312	10	0.04–0.84	1.07	.0029	.054	1.00–1.14
Agreeableness	4,288	10	−0.25	.308	9	−0.64–0.15	0.97	.0021	.046	0.91–1.03
Conscientiousness	4,565	12	0.53	.228	17	0.24–0.82	1.08	.0015	.039	1.03–1.13
Average			*0.51*	*.266*			*1.11*		*.053*	
Non-motivated incumbents										
Emotional stability	1,179	8	0.14	.115	67	−0.01–0.29	1.02	.0039	.062	0.94–1.10
Extraversion	1,331	9	0.18	.370	17	−0.29–0.66	1.04	.0048	.069	0.95–1.13
Openness	1,264	9	0.18	.510	10	−0.48–0.53	1.04	.0070	.083	0.93–1.15
Agreeableness	1,179	8	−0.05	.238	33	−0.35–0.25	1.00	.0021	.046	0.94–1.06
Conscientiousness	1,331	9	0.14	.126	63	−0.02–0.31	1.03	.0014	.038	0.98–1.08
Average			*0.12*	*.272*			*1.03*		*.060*	

Note: N = total sample; K = number of studies; d = sample size weighted mean effect size; SD_d = standard deviation in effect size remaining after sampling error variance is removed; %VE = % of observed variance accounted for sampling error; 80%CI_d = 80% credibility interval of effect sizes; FR = weighted-sample mean; SD_{fr} = standard deviation of FR; 80%CI_{fr} = 80% confidence interval for FR.

Table 3. Effect size (Cohen's *d*) and faking ratio (FR) of incumbents and applicants for the facets of the Big Five personality dimensions

Variable	N	K	d	SD_d	%VE	80%CI_d	FR	S^2_{fr}	SD_{fr}	80%CI_{fr}
Incumbents – all										
Emotional stability	962	8	−0.24	.217	42	−0.02–0.52	1.06	.0026	.051	0.99–1.13
Extraversion	962	8	0.48	.000	100	0.48	1.10	.0005	.022	1.07–1.13
Openness	3,713	10	0.36	.172	27	0.14–0.58	1.07	.0023	.048	1.01–1.13
Agreeableness	962	8	−0.08	.000	100	–0.08	0.98	.0002	.016	0.96–1.00
Conscientiousness	3,628	10	0.42	.119	44	0.26–0.57	1.08	.0008	.028	1.04–1.12
Average			*0.19*	*.102*			*1.06*		*.030*	
Applicants										
Emotional stability	28,943	8	0.88	.075	18	0.78–0.98	1.32	.0030	.055	1.08–1.22
Extraversion	28,943	8	0.52	.090	12	0.41–0.64	1.11	.0001	.011	1.10–1.13
Openness	29,261	9	0.29	.085	15	0.18–0.40	0.97	.0034	.058	0.90–1.04
Agreeableness	28,943	8	0.36	.099	10	0.24–0.49	1.07	.0004	.021	1.04–1.10
Conscientiousness	29,346	10	0.93	.142	7	0.74–1.11	1.18	.0006	.025	1.15–1.21
Average			*0.60*	*.098*			*1.13*		*.034*	

Note: N = total sample; K = number of studies; d =sample size weighted mean effect size; SD_d = standard deviation in effect size remaining after sampling error variance is removed; %VE = % of observed variance accounted for sampling error; 80%CI_d = 80% credibility interval of effect sizes; FR = weighted-sample mean; SD_{fr} = standard deviation of FR; 80%CI_{fr} = 80% confidence interval for FR.

found to conduct the meta-analysis of the facets of the Big Five for the within-subjects designs. The results for the incumbent group showed positive *d*-values for emotional stability, extraversion, openness, and conscientiousness, but the magnitude of the *d*-values was remarkably smaller than the values for the applicants. With regard to the applicants, the results were positive for the five personality factors, with values ranging from 0.29 to 0.93. With regard to FR, it varied from 0.98 to 1.10, with the largest value for extroversion, in the case of incumbents. For applicants, FR ranges from 0.97 to 1.32, with the largest value being for emotional stability. These results support *hypothesis 1a*, as the applicants showed larger FRs than the incumbents and the difference between the average ratio is significant ($p < .05$, one-tail). In the case of facets, not enough data were obtained to conduct the meta-analysis of the facets of the Big Five for the within-subjects designs.

Taken together, the results of the meta-analytic findings reported in Tables 1–3 fully support *Hypotheses 1a* and *1b* of the theoretical model, as applicants showed larger FR than motivated incumbents, with both groups showing remarkably larger FR than the nonmotivated incumbents and the normative group, while these latter two showed similar FRs (FR average = 1.03). Additionally, as was expected, the within-subjects design showed the largest FRs, which agree with Viswesvaran and Ones' (1999) findings.

7.2. *Meta-analysis of the coefficient of homogeneity* u *of the Big Five and their facets across samples*

The results of the meta-analysis of the coefficient of homogeneity *u* for the Big Five personality dimensions appear in Table 4. The results for the incumbents showed an average of 0.97, with the *u*-values ranging from 0.91 to

1.02 for emotional stability and openness, respectively. This finding suggests that the samples of incumbents are very slightly restricted compared with the normative groups. The results for the samples of applicants showed an important range restriction, as the average was 0.82, with *u*-values ranging from 0.78 to 0.89 for openness to agreeableness, respectively. The results for emotional stability and conscientiousness were 0.81 and 0.80, respectively. These two values are especially relevant as these two dimensions were the only ones that showed validity generalization across occupations (Barrick, Mount, & Judge, 2001; Salgado, 1997). The *u*-values for the applicants are also important because they are remarkably lower than the *u*-values used in previous meta-analyses of the validity of the Big Five. For example, the meta-analyses by Barrick and Mount (1991), Hurtz and Donovan (2000), Salgado (1997) used values of 0.94, 0.92, and 0.94, respectively. The results for the within-subject designs showed results similar to the applicants, but slightly less restricted (average 0.85). Consequently, the results reported in Table 4 fully support *hypothesis 2a*, which suggests that faking results in a homogenization of the groups.

Table 4. Coefficient of homogeneity *u* (range restriction value) of the three types of samples for the Big Five personality dimensions

Variable	*N*	*K*	*u*	S_u^2	SD_u	80%CoI
Incumbents – all						
Emotional stability	5,566	18	0.91	.0047	.069	0.82–1.00
Extraversion	5,897	20	0.98	.0052	.052	0.91–1.05
Openness	5,518	19	1.02	.0111	.111	0.88–1.16
Agreeableness	5,240	17	0.97	.0121	.121	0.82–1.22
Conscientiousness	5,769	20	0.96	.0106	.106	0.82–1.10
Average			*0.97*		*.092*	
Applicants						
Emotional stability	32,599	11	0.81	.0026	.051	0.75–0.87
Extraversion	32,599	11	0.82	.0024	.049	0.76–0.88
Openness	32,917	12	0.78	.0040	.063	0.70–0.86
Agreeableness	31,203	10	0.89	.0001	.012	0.88–0.90
Conscientiousness	33,002	13	0.80	.0037	.061	0.72–0.88
Average			*0.82*		*.047*	
Within-subject design						
Emotional stability	427	4	0.86	.0039	.063	0.78–0.94
Extraversion	607	5	0.79	.0303	.174	0.57–1.01
Openness	607	5	0.84	.0090	.096	0.72–0.96
Agreeableness	607	5	0.85	.0094	.098	0.73–0.97
Conscientiousness	607	5	0.92	.0037	.061	0.85–0.99
Average			*0.85*		*.098*	

Note: *N* = total sample; *K* = number of studies; *u* = weighted-sample homogeneity coefficient; S_u^2 = weighted-sample observed variance of *u*; SD_u = standard deviation of *u*; 80%CoI = 80% confidence interval for *u*.

Table 5 reports the meta-analysis of the coefficients of homogeneity for the motivated and nonmotivated to fake incumbents. As can be seen, the results for the nonmotivated group were close to 1 or slightly larger than 1 in all cases, with an average of 1.03. This means that the nonmotivated incumbents are similar to the normative groups or even that the range is slightly enhanced (heterogeneous). However, the results for the motivated incumbents showed a restricted range, which supports *hypothesis 2a*. The average was 0.95 for the Big Five as a whole, which is practically identical to the value used in the seminal meta-analyses of the validity of the Big Five (e.g., Barrick & Mount, 1991; Salgado, 1997). This last finding is relevant because practically all validity studies to date have been conducted with incumbents. The comparison of the *u*-values for the motivated incumbents and the applicant samples suggests that the applicants are more restricted in range than the incumbents. As the meta-analyses of the criterion validity of the Big Five used a *u*-value larger than the *u* corresponding to the applicants taking part in the selection process, the current results suggest than there may have been significant underestimation of the operational validity of the Big Five in these meta-analyses.

Table 5. Coefficient of homogeneity *u* (range restriction value) of motivated and non-motivated incumbents for the Big Five personality dimensions

Variable	*N*	*K*	*u*	S_u^2	SD_u	80%CoI
Motivated incumbents						
Emotional stability	4,161	9	0.88	.0021	.046	0.82–0.94
Extraversion	4,340	10	0.96	.0028	.053	0.89–1.03
Openness	4,250	10	1.02	.0114	.107	0.88–1.16
Agreeableness	4,161	9	0.95	.0040	.063	0.87–1.03
Conscientiousness	4,340	10	0.92	.0045	.067	0.83–1.01
Average			*0.95*		*.067*	
Non-motivated incumbents						
Emotional stability	1,405	9	0.98	.0049	.070	0.89–1.07
Extraversion	1,557	10	1.02	.0099	.099	0.89–1.15
Openness	1,264	9	1.01	.0105	.102	0.88–1.14
Agreeableness	1,079	9	1.06	.0347	.186	0.82–1.30
Conscientiousness	1,429	10	1.08	.0105	.103	0.95–1.21
Average			*1.03*		*.112*	

Note: *N* = total sample; *K* = number of studies; *u* = weighted-sample homogeneity coefficient; S_u^2 = weighted-sample observed variance of *u*; SD_u = standard deviation of *u*; 80%CoI = 80% confidence interval for *u*.

Table 6. Coefficient of homogeneity u (range restriction value) of applicants and incumbents for the facets of the Big Five personality dimensions

Variable	N	K	u	S_u^2	SD_u	80%CoI
Incumbents – All						
Emotional Stability	921	7	0.84	.0032	.057	0.77–0.91
Extraversion	921	7	0.88	.0028	.053	0.81–0.95
Openness	3,579	8	1.00	.0021	.046	0.96–1.06
Agreeableness	921	7	0.94	.0025	.050	0.88–1.00
Conscientiousness	3,579	8	0.95	.0022	.046	0.89–1.01
Average			*0.92*		*.050*	
Applicants						
Emotional Stability	28,943	8	0.82	.0011	.033	0.78–0.86
Extraversion	28,943	8	0.85	.0006	.025	0.82–0.88
Openness	29,261	9	0.88	.0001	.012	0.86–0.90
Agreeableness	28,943	8	0.91	.0001	.010	0.90–0.92
Conscientiousness	29,261	9	0.81	.0010	.031	0.77–0.85
Average			*0.85*		*.022*	

Note: N = total sample; K = number of studies; u = weighted-sample homogeneity coefficient; S_u^2 = weighted-sample observed variance of u; SD_u = standard deviation of u; 80%CoI = 80% confidence interval for u.

Table 7. Reliability (internal consistency) of the Big Five personality dimensions for faking and non-faking samples

Variable	N	K	α	S_α^2	SD_α	80%CoI
Non-faking						
Emotional stability	3,743	8	0.91	.0004	.020	0.88–0.94
Extraversion	3,743	8	0.88	.0027	.052	0.81–0.95
Openness	3,922	9	0.87	.0009	.030	0.83–0.91
Agreeableness	3,743	8	0.88	.0016	.040	0.83–0.93
Conscientiousness	3,922	9	0.91	.0002	.016	0.89–0.93
Average			*0.89*		*.032*	
Faking						
Emotional stability	2,753	5	0.87	.0010	.031	0.83–0.81
Extraversion	2,753	5	0.83	.0024	.049	0.77–0.89
Openness	3,073	6	0.85	.0026	.051	0.79–0.91
Agreeableness	2,753	5	0.86	.0028	.053	0.79–0.93
Conscientiousness	3,154	6	0.89	.0022	.047	0.83–0.95
Average			*0.86*		*.046*	

Note: N = total sample; K = number of studies; α = weighted-sample homogeneity coefficient; S_α^2 = weighted-sample observed variance of α; SD_α = standard deviation of α; 80%CoI = 80% confidence interval for α.

The meta-analysis of the u values for the facets of the Big Five appears in Table 6. In the case of the incumbent samples, the u-values range from 0.84 to 1, with an average of 0.92. Therefore, the incumbent samples, on average, showed slight range restriction, which corresponds with the u value used in the seminal meta-analyses. In the case of the samples of applicants, the u values ranged from 0.81 to 0.91, with an average of 0.85. Therefore, in the case of the facets, the samples of applicants are also more restricted than the samples of incumbents, as suggested by *hypothesis 2a*.

The results of this second set of meta-analyses are especially relevant for the theoretical model of faking, as they fully support the idea that motivation to fake (and actually faking) produces homogenization of the score distributions. This is a key aspect of the theoretical model and its confirmation was important. These findings also demonstrated that faking reduces the range of scores, especially for the personality dimensions.

7.3. Meta-analysis of the reliability coefficients for the Big Five and their facets under faking and nonfaking conditions

Hypothesis 3 states that the reliability of the personality scales is reduced as a consequence of faking. The results of the meta-analyses carried out to test this hypothesis appear in Tables 7 and 8. As can be seen in Table 7, the reliability under conditions that motivate to fake is smaller for the five dimensions than the reliability in nonmotivating conditions. The average reliability is 0.86 under faking conditions and 0.89 under nonmotivating to fake conditions. Globally, the reliability under faking conditions is 3% smaller than in nonmotivating-to fake conditions.

Table 8. Reliability (internal consistency) of the facets of the Big Five personality dimensions for faking and non-faking samples

Variable	N	K	α	S_α^2	SD_α	80%CoI
Non-faking						
Emotional stability	3,922	31	0.74	.0010	.032	0.70–0.78
Extraversion	3,922	32	0.71	.0005	.023	0.68–0.74
Openness	4,275	44	0.71	.0010	.031	0.67–0.75
Agreeableness	4,102	33	0.70	.0018	.042	0.65–0.75
Conscientiousness	4,362	40	0.71	.0023	.048	0.65–0.77
Average			*0.72*		*.035*	
Faking						
Emotional stability	6,325	19	0.67	.0007	.027	0.64–0.70
Extraversion	6,325	20	0.64	.0004	.026	0.61–0.67
Openness	6,819	26	0.66	.0022	.046	0.60–0.72
Agreeableness	6,505	21	0.63	.0024	.049	0.57–0.69
Conscientiousness	6,906	28	0.65	.0018	.043	0.60–0.70
Average			*0.65*		*.038*	

Note: N = total sample; K = number of studies; α = weighted-sample homogeneity coefficient; S_α^2 = weighted-sample observed variance of α; SD_α = standard deviation of α; 80%CoI = 80% confidence interval for α.

This finding agrees with the well-known result that range restriction produces a slightly reduction of the size of the reliability coefficients (Guilford, 1954; Gulliksen, 1950). In fact, for the same u-value, the percentage of reduction of the reliability coefficients is proportionally smaller than the respective percentage of reduction of the validity coefficients. However, it is important to take into account that the effects on reliability and validity are independent and cumulative.

The meta-analytic results for the reliability coefficients of the facets of the Big Five appear in Table 8. As can be seen, the reliability average is 0.72 under nonfaking conditions and 0.65 under faking conditions. Globally, the increment of the measurement error is 10%. This suggests that effect of the range restriction due to faking is greater for the reliability of facets than for the reliability of factors, even when the range restriction was slightly larger for the factors than for the facets. An additional finding is that the reliability of the scales is lower that the minimal requirement of reliability suggested by Nunnally (1978) and that many researchers use as a typical cut-off-point to decide if a scale is reliable.

As a whole, the results reported in Tables 7 and 8 confirmed *Hypothesis 3* and provide empirical evidence supporting the predictions of the theoretical model. They are also relevant in connection with Heggestad et al.'s (2006) finding that an alpha coefficient overestimates the reliability of faked scales. If, on one hand, faking reduces the size of the alpha estimates, according to the results of Tables 7 and 8, and, on the other hand, the alpha overestimates the reliability, the suggestion is that the reliability under faking conditions may be even worse than the current alpha estimates suggest.

8. Discussion

The theoretical model presented here suggests that faking is a phenomenon of homogenization of distribution of scores through the intentional distortion of responses to assessment procedures. Individuals adapt their responses with the objective of managing the impressions they cause and of presenting themselves in a more effective way for their purposes (e.g., getting a job). It directly affects the mean, standard deviation, reliability, and validity of measures by a double mechanism. On the one hand, faking increases means (in the case of faking good) or reduces means (in the case of faking bad). On the other hand, faking reduces standard deviations (causing range restriction), reliability, and validity coefficients. Furthermore, faking indirectly affects the correlation of the faked variable with other variables, as well as the mean, standard deviation, and reliability estimates of the correlated variables. This last result is an effect of the indirect range restriction of the variables. Pearson (1903) mathematically demonstrated over a hundred years ago that the direct range restriction of X is always accompanied by the indirect restriction of the mean, and standard deviation of Y.

The model makes two predictions. The first is that the *d*-values and the ratio between the mean of a motivated-to-fake group and the mean of a nonmotivated-to fake group will be greater than 1 (in the case of faking good). This ratio was named FR. The second prediction is that the ratio between the standard deviation of the motivated-to-fake group and the nonmotivated-to-fake group will be smaller than 1. This second ratio is the coefficient of homogeneity u.

Three series of meta-analyses were carried out to test five hypotheses and the predictions derived from this model for the case of the Big Five personality dimensions and their facets. The first series confirmed that the *d*-values were positive and that the FR was larger than 1 for motivated-to-fake incumbents, applicants, and experimental subjects. It was also demonstrated that if the motivation-to-fake increases, the ratio increases, as the model suggests. The second series of meta-analyses examined one key issue of the model, that faking produces homogeneity in the distribution of scores. The coefficient of homogeneity u was smaller than 1 in all cases, as predicted. Additionally, if the motivation-to-fake increases, the u ratio decreases, as the model suggests. Finally, the third series of meta-analyses examined the effects of faking on the reliability coefficients. The model suggests that reliability of the scales is larger for the nonmotivated-to-fake groups, and the results supported this hypothesis. Furthermore, it was found that faking is even more 'dangerous' for the reliability of the facets than for the personality dimensions. This last finding was not anticipated but is important from an applied point of view as it suggests that scales with a large number of items should be used, as they are more robust against the effects of faking on measurement error. This last finding may be one of the reasons why facets show smaller criterion-oriented validity than the Big Five dimensions for predicting job performance (Salgado, Anderson, & Tauriz, 2015) due to the attenuation of validity since the reliability of facets is smaller than the reliability of the Big Five.

Some meta-analyses had demonstrated the effect size of the faking on the mean of personality inventories (Birkeland et al., 2006; Ones et al., 1996; Viswesvaran & Ones, 1999). They found that incumbents showed larger mean than nonincumbent samples. However, the study of the effects of faking on the variability of the distributions of scores has been neglected until now. According to the theoretical model, the effect of faking on the variability is another key aspect of the negative consequences of faking.

The results of these meta-analyses also showed that personality scales responded to by applicants show a pattern of means, standard deviations, and reliabilities that is similar to the pattern found for individuals in experimental studies in which within-subject designs are used. In fact, applicants actually appear to be more homogenous groups than experimental groups, in term of their respective standard deviation. Therefore, faking in real-life situations (e.g., personnel selection) and in artificial (laboratory) situations appears to produce the same effects. Additionally, motivated-to-fake incumbents show essentially the same pattern of responses, although faking appears to be less intense for them.

8.1. Implications for research and practice on personality at work

The range restriction produced by faking necessarily affects the size of reliability coefficients, the criterion validity, and the construct validity of personality inventories in all personality dimensions in samples of applicants, but its effects are not perceived or are hidden when they are examined in samples of nonmotivated-to-fake individuals (e.g., nonmotivated-to fake incumbents). Consequently, the present findings have implications for meta-analyses of the criterion-oriented validity of personality measures as many previous meta-analyses (e.g., Barrick & Mount, 1991; Hurtz & Donovan, 2000; Salgado, 1997) used RR distributions based on incumbent samples. All the *u* values reported in previous meta-analyses were based on the incumbent *u*, because virtually all the criterion-related validity studies in personnel selection were carried out with incumbents (Schmidt Oh, & Le, 2006). They are very similar to the *u* values reported in the present study for incumbents as a whole and for motivated incumbents in particular, ranging from 0.92 to 0.94. The current findings show that the distributions based on incumbents strongly underestimate the actual range restriction if they are used for samples of applicants. Therefore, future meta-analyses should be take into account the difference in range restriction distributions and to use the appropriate for each single case.,

Another implication of the results is that, as a consequence of faking, the mechanism of range restriction in personality measures appears to be different from the mechanism that operates for GMA and cognitive ability tests. Whereas range restriction in GMA tests is a product of the selection decision, the range restriction in personality appears to be mainly a product of faking. In other words, GMA range restriction occurs after the decision but personality range restriction occurs prior to the decisions. For this reason, the range restriction of GMA in incumbents is larger than in applicants, but the range restriction of personality measures tends to disappear in incumbents, while it is large in applicants. In other words, the mechanism of range restriction operates in GMA for incumbents but it operates in personality for applicants. This is an important distinction because the selection is made from the applicants, which are the restricted group in personality.

The theoretical model and the findings reported here also have implications for the mechanical correction of individual scores. These methods would produce a larger range restriction, because they change a high score to a smaller one but the lower scores (but faked) are not corrected. Subsequently, this correction method would have perverse effects on reliability and validity as the reduction of scores in individuals scoring high on social desirability scales is not accompanied by a similar reduction in the scores of the individuals scoring low and medium on the *SD* scales. Therefore, the standard deviations produced by the mechanical corrections would be even smaller than the observed standard deviations in the applicant sample, resulting in an increased range restriction, which in turn would affect reliability and validity coefficients.

The theoretical model of faking presented here suggests that an appropriate method for controlling the effects of faking will be one that prevents or attenuates faking, by deflating the mean and increasing the standard deviation, so that the mean and the standard deviation of samples of motivated-to-fake individuals are similar to the mean and standard deviation of the normative group. Among the current methods, warning candidates not to fake, together with norms based on applicant samples appear to be the preferable methods.

The model of faking effects and the results of the present research have also implication for the construct validity of personality measures. With regard to construct validity, it is important to take into account the fact that faking effects can be either univariate or sequential depending on the number of personality variables affected. If faking is univariate, it produces an underestimation of the correlations between the faked variables and other related variables. If faking is sequential, it can produce extra factors, turning an orthogonal structure into an oblique one (Thurstone, 1945, 1965).

With regard to this, Ellingson et al. (2001) posed an interesting and relevant question, How can score structure remain similar across two groups while mean scores differ? They offered two potential explanations. The first is that faking would add a constant to the scores but the addition of the constant would not modify the correlations calculated among the scales. However, this explanation does not reflect the mechanism through which faking modifies observed scores. This explanation was also suggested by Cronbach (1946) when he wrote that response sets (e.g., social desirability) might also be compared with constant errors in psychophysics: the error may be 'constant' for the individual. The second explanation is that faking actually reflects true variance, even if only to some degree, and, consequently, it suggests that there are real individual differences in faking.

The theoretical model of faking and the findings of the current study offer an alternative answer to Ellingson et al.'s question. The structure remained similar because the range restriction was univariate in their case. In other words, faking would have directly affected a single variable (and the others indirectly) in their study. The changes to the mean do not affect the structure. The changes in the structure are mainly due to the magnitude of the changes in the variances and covariances of the variables analyzed.

Additionally, the results of this study are also relevant for within-subject designs. The effects of faking on the range restriction in samples of applicants are slightly larger than the effects observed in experimental studies using within-subject designs, an effect unnoticed until now.

Based on the findings that this design type showed similar (but conservative) results to the one found in applicant samples, within-subject designs may be useful for developing predictive indexes of faking and for validating social desirability scales.

8.2. Implications for human resource assessment

The faking model and the empirical findings observed in this research also have at least two relevant implications for the practice of HR assessment and personnel selection. As was predicted and has been demonstrated, faking produces larger homogeneity of applicant samples than in incumbent samples. This effect is important in connection with the use of norms based on samples of incumbents and nonmotivated-to-fake individuals. These norms based on incumbents would have lower means and larger *SD*s than the corresponding mean and *SD* of the applicant sample. Therefore, the scores of the applicants can be seriously biased if the incumbent norms are applied. This problem of the larger homogeneity of applicant also has important implications for the typical practice of setting cut scores based on incumbents and then using that cutting score for applicants. Basically, this practice will result in larger pass rates. Therefore, practitioners should be aware of the type of samples used for the development of norms. With regard to these two problems, the suggestion is to use norms based on samples of applicants.

A second implication of the findings refers to the use of measures of personality facets for personnel selection purposes. The reliability of the facets is dramatically affected by faking, to the point that on average the reliability was 0.65 under faking conditions, which is an unacceptable figure for practical purposes. Therefore, if practitioners use measures of facets for making personnel decisions, they should examine the reliability of these measures in every specific case and to estimate the standard error of measurement for that sample.

8.3. Limitations of the study and future research

A word about the advantages and disadvantages of using the NEO-PI-R appears to be needed. The advantage of using the NEO PI-R to control the effects of the variability of inventories on the construct validity of the Big Five is accompanied by the limitation of the generalizability of the results. It is possible that the present results may only be valid for the NEO PI-R and, therefore, additional research using a variety of inventories would be desirable. Nevertheless, the results of some studies suggest that the present findings are not specific to the NEO PI-R. For example, O'Brien and LaHuis (2011) found that applicants showed larger means and smaller *SD* than incumbents in two large samples ($n = 1{,}509$ and $n = 1{,}568$) in a study using the 16PF.

Another limitation of this study is that the RR estimates found here are applicable only to single-stimulus (SS) personality inventories. At present, they are not applicable to forced-choice (FC) inventories (Salgado & Tauriz, 2014; Salgado, Moscoso et al., 2015). Although FC inventories may be affected by faking, the effect may be different both for the mean and the *SD*. Future studies should examine this issue. A third limitation is that the effects of faking were examined for the Cronbach's alpha coefficient of reliability. Probably, the effects of faking apply to other reliability coefficients such as the test–retest coefficient, the interrater coefficient, and the CES (Salgado, 2015; Salgado, Moscoso, & Anderson, 2016; Salgado & Moscoso, 1996; Schmidt, Le, & Ilies, 2003), but additional research is required.

Future studies should also be carried out to investigate whether the prediction that univariate and sequential faking have differential consequences on the structure of personality inventories. If the predictions of the model are correct, univariate faking should produce reduced loadings but essentially the same structure, while, if faking is sequential, additional factors can be expected, depending on the number of variables affected by faking and the number of variables included in the factor analyses. In addition, future studies should examine the prediction that the validity of personality inventories is negatively affected by faking, using designs that directly control faking effects. For example, employees could take a personality inventory under honest and faking instructions and their results could be correlated with job performance.

9. Conclusion

This article suggests a theoretical model that conceptualizes faking as a phenomenon of homogenization of the scores of assessment procedures through a double mechanism that operates on the mean and the standard deviation of measures. This double mechanism has negative consequences for the reliability and validity of measures. The empirical findings reported support some predictions of the faking model (e.g., larger means, smaller *SD*, reduced reliability). However, this is just the beginning and future research should be carried out for testing additional predictions.

Acknowledgement

The research reported in this manuscript was supported by Grant PSI2014-56615-P from the Spanish Ministry of Economy and Competitiveness.

Note

1. I appreciate Frank Schmidt's clarification on this important distinction.

References

Articles included in the meta-analysis are preceded by *

Anderson, N., & Sleap, S. (2004). An evaluation of gender differences on the Belbin team role self-perception inventory. *Journal of Occupational and Organizational Psychology*, 77, 429–437.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211.

Bangerter, A., Roiulin, N., & Köning, C. J. (2012). Personnel selection as a signaling game. *Journal of Applied Psychology*, 97, 719–739.

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. DOI: 10.1111/1468-2389.00160.

Birkeland, S., Manson, T., Ksiamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317–335.

*Black, J. (2000). *Personality testing and police selection: Utility of the 'big five'*. New Zealand Journal of Psychology, 29, 2–9.

Brogden, H. E. (1944). On the estimation of the changes in correlation and regression constants due to selection on a single given variable. *Journal of Educational Psychology*, 25, 484–492.

Burt, C. (1943). Validating tests for personnel selection. *British Journal of Psychology*, 34, 1–19.

*Caldwell-Andrews, A., Baer, R. A., & Berry, D. T. R. (2000). Effects of response sets on NEO-PI-R scores and their relations to external criteria. *Journal of Personality Assessment*, 74, 472–488.

*Cellar, D. F., Miller, M. L., Doverspike, D. D., & Klawsky, J. D. (1996). Comparison of factor structures and criterion-related validity coefficients of two measures of personality based on the five factor model. *Journal of Applied Psychology*, 81, 694–704.

*Chibnall, J. T., & Detrick, P. (2003). The NEO-PI-R, inwald personality inventory, and MMPI-2 in the prediction of police academy performance: A case for incremental validity. *American Journal of Criminal Justice*, 27, 233–247.

Christiansen, N. D., Goffin, R. D., Johnson, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology*, 47, 847–860.

Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations. *International Journal of Selection and Assessment*, 11, 17–29.

*Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

*Costa, P. T., & McCrae, R. R. (1999). *Manual profesional*. NEO–PI–R Inventario de Personalidad NEO Revisado. Madrid: TEA ediciones.

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6, 475–494.

*Denis, P. L., & Gilbert, F. (2012). The effect of time constraints and personality facets on general cognitive ability (GCA) assessment. *Personality and Individual Differences*, 52, 541–545.

*Denis, P. L., Martin, D., & Guindan, C. (2010). Exploring the capacity of NEO PI-R facets to predict job performance in two French-canadian samples. *International Journal of Selection and Assessment*, 18, 201–207.

*Detrick, P., Chibnall, J. T., & Call, C. (2010). Demand effects on positive response distortion by police officer applicants on the revised NEO personality inventory. *Journal of Personality Assessment*, 92, 410–415.

*Detrick, P., Chibnall, J. T., & Luebbert, M. C. (2004). The revised NEO personality inventory as predictor of police academy performance. *Criminal, Justice and Behavior*, 31, 676–694.

Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996). *The validity of non-cognitive measures decays when applicants fake*. Paper presented at the annual meeting of the *Academy of Management, Cincinnati, OH*,

Dunlap, W. P., & Silver, N. C. (1986). Confidence intervals and standard errors for ratios of normal variables. *Behavior Research Methods, Instruments, & Computers*, 18, 469–471.

Ellingson, J. E., 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. DOI: 10.1037/0021-9010.84.2.155.

Ellingson, 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. DOI: 10.1037/0021-9010.86.1.122.

Ellis, A. (1946). The validity of personality questionnaires. *Psychological Bulletin*, 43, 385–440.

Ferguson, E., & Bibby, P. A. (2012). Openness to experience and all-cause mortality: A meta-analysis and *r* equivalent from risk ratios and odds ratios. *British Journal of Health Psychology*, 17, 85–102.

Fieller, E. C. (1932). The distribution of the index of a normal bivariate population. *Biometrika*, 24, 428–440.

Frei, R. L., Griffith, R. L., McDaniel, M. A., Snell, A. F., Hamill, L. S., Wheeler, J. K., et al. (1997). *Building a better mousetrap: Using CART to detect response distortion*. Paper presented at the Twelfth Annual Conference of the Society of Industrial and Organizational Psychology, Inc. St. Louis.

Furnham, A. (1986). Response bias, social desirability, and dissimulation. *Personality and Individual Differences*, 7, 385–400.

Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology: Research & Reviews*, 9, 46–55.

Geary, R. C. (1930). The frequency distribution of the quotient of two normal variates. *Journal of the Royal Statistical Society*, 93, 442–446.

Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology*, 50, 151–160.

*Grice, R. L., & Katz, L. C. (2007a). *Personality profiles of U.S. Army initial entry rotary wing students versus career aviators*.

Technical report 1185. *United States Army Research Institute for the Behavioral and Social Sciences.*

*Grice, R. L., & Katz, L. C. (2007b). *Personality profiles of U.S. Army initial entry rotary wing students versus career aviators*. Technical report 1208. *United States Army Research Institute for the Behavioral and Social Sciences.*

Grieve, R., & McSwiggan, (2014). Predicting intentions to fake in psychological testing: Which normative beliefs are important? *Revista De Psicología Del Trabajo Y De Las Organizaciones*, 30, 23–28.

*Griffin, B., & Hesketh, B. (2004). Why openness to experience is not a good predictor of performance. *International Journal of Selection and Assessment*, 12, 243–251.

*Griffin, B., Hesketh, B., & Grayson, D. (2004). Applicants faking good: Evidence of item bias in the NEO-PI-R. *Personality and Individual Differences*, 36, 1545–1558.

Griffin, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? an examination of the frequency of applicant faking behavior. *Personnel Review*, 36, 341–345.

Griffith, R. L., Lee, L. M., Peterson, M. H., & Zickar, M. J. (2011). First dates and little white lies: A trait contract classification theory of applicant faking behavior. *Human Performance*, 24, 338–357.

Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley

*Hart, P. M. (1999). Predicting employee life satisfaction: A coherent model of personality work and nonwork experiences, and domain satisfaction. *Journal of Applied Psychology*, 84, 564–584.

Hedges, L., Gurevicth, J., & Curtis, P. S. (1999). The meta-analysis of response ratios in experimental ecology. *Ecology*, 80, 1150–1156.

Heggestad, E. D., George, E., & Reeve, C. L. (2006). Trasient error in personality scores: Considering honest ad faked responses. *Personality and Individual Differences*, 40, 1201–1211.

Hogan, R. T., & Hogan, J. (1995). *Manual for the Hogan personality inventory* (2nd ed.) Tulsa, OK: Hogan Assessments.

Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11, 209–244.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis. Correcting error and bias in research findings* (2nd ed.) Newbury Park, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and direct range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91, 594–612.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology, 85*, 869–879. DOI: 10.1037/0021-9010.85.6.869

*Jackson, C. J., & Corr, P. J. (1998). Personality–performance correlations at work: Individual and aggregate levels of analyses. *Personality and Individual Differences*, 24, 815–820.

*Juhel, J., Brunot, S., & Zapata, G. (2012). Response distortion on the NEO-PI-R among candidates taking the entrance examination to the national school of civil aviation (ENAC-france). *Psychology*, 3, 393–398.

Kelley, T. L. (1921). The reliability of test scores. *Journal of Educational Research*, 3, 370–379.

Kelley, T. L. (1947). *Fundamentals of statistics*. Cambridge, MA: Harvard University Press.

Lajeunesse, M. J. (2011). On the meta-analysis of response ratios for studies with correlated and multi-groups designs. *Ecology*, 92, 2049–2055.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulas. *Proceedings of the Royal Society of Edinburgh, Section A*, 62, Part I, 20–30.

Ledermann, W. (1938). Note on Professor Godfrey H. Thomson's article "the influence of univariate selection on factorial analysis of ability". *British Journal of Psychology*, 29, 69–73.

*Lee, Y.-T., Stettler, A., & Antonakis, J. (2011). Incremental validity and indirect effect of ethical development on work performance. *Personality and Individual Differences*, 50, 1110–1115.

Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation o fan interview faking behavior scale. *Journal of Applied Psychology*, 92, 1638–1656.

*Major, D. A., Turner, J. E., & Fletcher, T. D. (2006). Linking proactive personality and the big five to motivation to learn and development activity. *Journal of Applied Psychology*, 91, 927–933.

Marcus, B. (2009). 'Faking' from the applicant's perspective: A theory of self-presentation in personnel selection settings. *International Journal of Selection and Assessment*, 17, 417–430.

*Marshall, M. B., De Fruyt, F., Rolland, J. P., & Bagby, R. M. (2005). Socially desirable responding and the factorial stability of the NEO-PI-R. *Psychological Assessment*, 17, 379–384.

McFarland, L. A., & Ryan, A. M. (2000). Variance of faking across noncognitive measures. *Journal of Applied Psychology*, 85, 812–821.

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the MMPI. *Journal of Applied Psychology*, 30, 525–564.

Michaelis, W., & Eysenck, H. J. (1971). The determination of personality inventory factor patterns and inter-correlations by changes in real life motivation. *Journal of Genetic Psychology*, 118, 223–234.

*Minbashian, A., Bright, J. E. H., & Bird, K. D. (2009). Complexity in the relationships among the subdimensions of extroversion and job performance in managerial occupations. *Journal of Occupational and Organizational Psychology*, 82, 537–549.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel contexts. *Personnel Psychology*, 60, 683–729.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Are we getting fooled again? coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, 60, 1029–1049.

*Moutafi, J., Furham, A., & Crump, J. (2006). What facets of openness and conscientiousness predict fluid intelligence score? *Learning and Individual Difference*, 16, 31–42.

Mueller-Hanson, R. A., Heggestad, E. D., & Thornton, G. C., III. (2006). Individual differences in impression management: An exploration of the psychological processes undelying faking. *Psychology Science*, 48, 288–312.

Murphy, K. R., & Dzieweczysnki, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, 18, 343–357.

*Neuman, G. A., & Kickul, J. R. (1998). Organizational citizenship behaviors: Achievement orientation and personality. *Journal of Business and Psychology*, 13, 263–279.
Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.) New York: McGraw-Hill.
O'Brien, E., & LaHuis, D. M. (2011). Do applicants and incumbents respond to personality items similarly? a comparison of dominance and ideal point response models. *International Journal of Selection and Assessment*, 19, 109–118.
Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, 11, 245–270.
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.
Otis, A. S. (1922). A method of inferring the change in a coefficient of correlation resulting from a change in the heterogeneity of the group. *Journal of Educational Psychology*, 13, 293–294.
*Pace, V. L., & Brannick, M. T. (2010). Improving prediction of work performance through frame-of-reference consistency: Empirical evidence using openness to experience. *International Journal of Selection and Assessment*, 18, 230–235.
Pearson, K. (1903). On the influence of natural selection on the variability and correlation of organs. *Philosophical Transaction of the Royal Society- Series A*, 200, 1–66.
*Raymark, P. H., & Tafero, T. L. (2009). Individual differences in the ability to fake on personality measures. *Human Performance*, 22, 86–103.
Ree, M. J., Carretta, T. R., Earles, J. A., & Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology*, 79, 298–301.
*Reid-Seiser, H. L., & Fritzsche, B. A. (2001). The usefulness of the NEO PI-R positive presentation management scale for detecting response distortion in employment contexts. *Personality and Individual Differences*, 31, 639–650.
*Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634–644.
Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology*, 75, 175–184.
Salgado, J. F. (1997). The five-factor model of personality and job performance in the European community. *Journal of Applied Psychology*, 82, 30–43. DOI: 10.1037/0021-9010.82.1.30
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. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, 76, 323–346.
Salgado, J. F. (2005). Personality and social desirability in organizational settings: Practical implications for work and organizational psychology. *Papeles Del Psicólogo*, 26, 115–128.
Salgado, J. F. (2015). Estimationg coefficients of equivalence and stability for job performance ratings: The importance of controlling for transient error on criterion measurement. *International Journal of Selection and Assesment*, 23, 37–44.
Salgado, J. F., Anderson, N., & Tauriz, G. (2015). The validity of ipsative and quasi-ipsative forced choice personality inventories for different occupational groups: A comprehensive meta-analysis. *Journal of Occupational and Organizational Psyhcology*, 88, 797–834.
Salgado, J. F., & Moscoso, S. (1996). Meta-analysis of interrater reliability of job performance ratings in validity studies of personnel selection. *Perceptual and Motor Skills*, 83, 1195–1201.
Salgado, J. F. & Moscoso, S. D. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology*, 11, 299–324.
Salgado, J. F., Moscoso, S., & Anderson, N. (2016). Corrections for criterion reliability in validity generalization: The consistency of Hermes, the utility of Midas. *Revista de Psicologia del Trabajo y de las Organizaciones*, 32, 17–23. doi:10.1016/j.rpto.2015.12.001
Salgado, J. F., Moscoso, S., Sanchez, J. I., Alonso, P., Choragwicka, B., & Berges, A. (2015). The validity of the five-factor model and their facets: The impact of performance measure and facet residualization on the bandwidth-fidelity dilemma. *European Journal of Work and Organizational Psychology*, 24, 325–349.
Salgado, J. F., & Tauriz, G. (2014). The five-factor model, forced-choice personality inventories, and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology*, 23, 3–30.
*Salters-Pedneault, K., Ruef, A. M., & Orr, S. P. (2010). Personality and psycholphysiological profiles of pólice officer and firefighter recruits. *Personality and Individual Differences*, 49, 210–215.
*Sanz, J., & García-Vera, M. P. (2009). New norms for the Spanish adaptation of the NEO personality inventory-revised (NEO PI-R): Reliability and normative data in volunteers from the general population. *Clinic and Health*, 20, 131–144.
Scarpello, V. G., Ledvinka, J., & Bergmann, T. J. (1995). *Human resource management*. Cincinnati, OH: South-Western College Publishing.
Schmidt, F. L. & Hunter, J. E. (2014). *Methods of meta-analysis* (3rd ed.) Newbury Park, CA: Sage.
Schmidt, F. L., & Le, H. (2014). *Software for the Hunter-Schmidt meta-analysis methods*. Iowa City, IA: Department of Management and Organizations, University of Iowa.
Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validity studies. *Journal of Applied Psychology*, 61, 473–485.
Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*, 8, 206–224.
Schmidt, F. L., Oh, I.-S., & Le, H. (2006). Increasing the accuracy of corrections for range restriction: Implications for selection procedure validities and other research practices. *Personnel Psychology*, 59, 281–305.
Schmit, M. J., & Ryan, A. M. (1993). The big five in personnel selection: Actor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78, 966–974.
Seisdedos, N. (1993). Personnel selection, questionnaires, and motivational distortion: An intelligent attitude of adaptation. In H. Schuler, J. L. Farr, & M. Smith (Eds.), *Personnel selection and assessment* (pp. 91–108). Hillsdale, NJ: Erlbaum.

*Shoss, M. K., & Strube, J. (2011). How do you fake a personality test? an investigation of cognitive models of impression-managed responding. *Organizational Behavior and Human Decision Processes*, 116, 163–171.

Smith, D. B., Hanges, P. J., & Dickson, M. W. (2001). Personnel selection and the five-factor model: Re-examining the effects of applicant's frame of reference. *Journal of Applied Psychology*, 86, 304–315. DOI: 10.1037/0021-9010.86.2.304

Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, 9, 219–242.

*Stewart, G. L. (1996). Reward structure as a moderator of the relationship between extraversion and sales performance. *Journal of Applied Psychology*, 81, 619–627.

*Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. *Journal of Applied Psychology*, 84, 959–968.

Tauriz, G. (2011). Falseamiento y validez de las medidas de personalidad en contextos académicos [aking and validity of personality measures in academia contexts]. *Revista e Psicología el Trabajo e as Organizaciones*, 27, 103–115.

Tett, R., Rothstein, M. G., & Jackson, D. J. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44, 703–742.

*Timmerman, T. A. (2006). Predicting turnover with broad and narrow personality traits. *International Journal of Selection and Assessment*, 14, 392–399.

Thomson, G. H. (1938). The influence of univariate selection on the factorial analysis of ability. *British Journal of Psychology*, 28, 451–459.

Thomson, G. H. (1939). *The factorial analysis of human ability*. New York: Houghton Mifflin.

Thomson, G. H. (1944). The applicability of Karl Pearson's selection formulae in follow-up experiments. *British Journal of Psychology: General Section*, 34, 105.

Thomson, G. H., & Ledermann, W. (1939). The influence of multivariate selection on the factorial analysis of ability. *British Journal of Psychology*, 29, 288–306.

Thorndike, R. L. (1947). *Personnel selection*. New York: Wiley.

Thurstone, L. L. (1935). *The vectors of the mind*. Chicago, IL: University of Chicago Press.

Thurstone, L. L. (1945). The effects of selection in factor analysis. *Psychometrika*, 10, 165–198.

Thurstone, L. L. (1965). *Multiple factor analysis* (2nd ed.) Chicago, IL: University of Chicago Press. (Original 1947).

*van der Linden, D., Te Nijenhuis, J., Cremers, M., & van de Ven, C. (2011). General factors of personality in six databases and a criterion-related validity study at the Netherlands armed forces. *International Journal of Selection and Assessment*, 19, 157–167.

Van Hooft, E. A., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, 97, 301–316. doi: 10.1037/a0025711

*Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology*, 90, 536–552.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analysis of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 60, 197–210.

*Winkerlspecht, C., Lewis, P., & Thomas, A. (2006). Potential effects of faking on the NEO-PI-R: Illingness and ability to fake changes who gets hired in simulated selection decisions. *Journal of Business and Psychology*, 21, 243–259.

*Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement*, 69, 548–565.

*Ziegler, M., Danay, E., Schölmerich, F., & Bühner, M. (2010). Predicting academic success with the Big 5 rated from different points of view: Self-rated, other rated and faked. *European Journal of Personality*, 24, 341–355.

Zickar, M. (2000). Modeling faking on personality tests. In D. Ilgen & C. L. Hulin (Eds.), *Computing meodeling behavioral processes in organization (95-108)*. Washington, DC: American Psychological Association.

Zickar, M., & Gibby, R. E. (2006). A history of faking and socially desirable responding on personality test. In R. L. Griffith & M. H. Petterson (Eds.), *A closer examination of applicants faking behavior*. (pp. 21–42). Charlotte, NC: IAP-Information Age Publishing.

Zickar, M., Rosee, J., & Levin, R. (1996). *Modelling the effects of faking on personality instruments*. Paper presented at the 11th Annual Meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.